



STIC EIC 2100 Search Request Form

133148

Today's Date:

09/21/04

What date would you like to use to limit the search?

Priority Date: 04/26/2000 Other:

Name TO, BAO QUOC

AU 2072 Examiner # 7889

Room # 4A42 Phone 305-1949

Serial # 09/559223

Format for Search Results (Circle One):

☒ PAPER ☐ DISK ☐ EMAIL

Where have you searched so far?

☒ USP ☐ DWPI ☒ EPO ☒ JPO ☐ ACM ☐ IBM TDB

☐ IEEE ☐ INSPEC ☐ SPI Other _____

Is this a "Fast & Focused" Search Request? (Circle One) ☒ YES ☐ NO

A "Fast & Focused" Search is completed in 2-3 hours (maximum). The search must be on a very specific topic and meet certain criteria. The criteria are posted in EIC2100 and on the EIC2100 NPL Web Page at <http://ptoweb/patents/stic/stic-tc2100.htm>.

What is the topic, novelty, motivation, utility, or other specific details defining the desired focus of this search? Please include the concepts, synonyms, keywords, acronyms, definitions, strategies, and anything else that helps to describe the topic. Please attach a copy of the abstract, background, brief summary, pertinent claims and any citations of relevant art you have found.

Please search for removing the located
term in the search

go to Subword
word

four

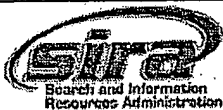
1+

STIC Searcher Cand Wong

Phone 305 9729

Date picked up 9-21

Date Completed 9-27-04



File 347:JAPIO Nov 1976-2004/May(Updated 040903)

(c) 2004 JPO & JAPIO

File 350:Derwent WPIX 1963-2004/UD,UM &UP=200459

(c) 2004 Thomson Derwent

| Set | Items | Description |
|-----|---------|--|
| S1 | 240219 | QUERY? OR QUERIE? ? OR SUBQUER? OR SEARCH? OR FETCH? OR RETRIEV? OR TEXTSEARCH? |
| S2 | 242131 | MATCH??? ? |
| S3 | 280897 | WORD? ? OR TERM? ? |
| S4 | 18386 | CHARACTERSTRING? OR CHARACTER? ?(2N)STRING? ? OR WORDSTEM? OR MORPHEME? OR WORDELEMENT? OR BASETERM? OR BASEWORD? OR LEX-EME? |
| S5 | 2047 | S3(2N) (STEM? ? OR ELEMENT? ? OR BASE OR BASES) |
| S6 | 428200 | SUFFIX? OR PREFIX? OR DERIVATI? OR AFFIX? OR POSTFIX? OR TRUNCAT? OR LEFTTRUNCAT? OR RIGHTTRUNCAT? |
| S7 | 1533926 | RANK? OR RATE OR RATES OR RATED OR RATING? OR SORT??? ? OR SCOR??? ? OR VALUATION? OR TALLY? OR TALLIE? ? OR WEIGH? |
| S8 | 43473 | S7(3N) (RESULT? OR HITLIST? OR REFERENCE? OR RETRIEV? OR HIT OR HITS OR OUTPUT? OR OUT()PUT? ? OR RESPONSE? ? OR ANSWER? ? OR REPLIE? ? OR REPLY?) |
| S9 | 7563 | S2:S4(3N)S7 |
| S10 | 26818 | S1:S2 AND S3:S5 |
| S11 | 486 | S10 AND S6 |
| S12 | 21 | S11 AND S8:S9 |
| S13 | 21 | IDPAT (sorted in duplicate/non-duplicate order) |
| S14 | 20 | IDPAT (primary/non-duplicate records only) |

? t14/9/4-6

14/9/4 (Item 4 from file: 350)

DIALOG(R)File 350:Derwent WPIX

(c) 2004 Thomson Derwent. All rts. reserv.

014765414 **Image available**

WPI Acc No: 2002-586118/200263

XRPX Acc No: N02-464928

Document search device assigns rank to searched document containing words which are extracted using word characteristic information containing suffix / prefix of independent word

Patent Assignee: CANON KK (CANO)

Number of Countries: 001 Number of Patents: 001

Patent Family:

| Patent No | Kind | Date | Applicat No | Kind | Date | Week |
|---------------|------|----------|---------------|------|----------|----------|
| JP 2002123545 | A | 20020426 | JP 2000317005 | A | 20001017 | 200263 B |

Priority Applications (No Type Date): JP 2000317005 A 20001017

Patent Details:

| Patent No | Kind | Lan Pg | Main IPC | Filing Notes |
|---------------|------|--------|-------------|--------------|
| JP 2002123545 | A | 12 | G06F-017/30 | |

Abstract (Basic): JP 2002123545 A

NOVELTY - A word characteristic information containing suffix / prefix of an independent word in a document is used to extract the word included in character row of the document. The other word characteristic information containing suffix / prefix of each extracted word is generated and the document containing extracted words is searched. A rank is assigned to searched document, based on characteristic information.

DETAILED DESCRIPTION - INDEPENDENT CLAIMS are included for the following:

(1) Document **search** method; and
(2) Recorded medium storing document **search** program.
USE - Document **search** device.
ADVANTAGE - The document is **searched** effectively.
DESCRIPTION OF DRAWING(S) - The figure shows the block diagram of document **search** device. (Drawing includes non-English language text).

pp; 12 DwgNo 1/13
Title Terms: DOCUMENT; **SEARCH** ; DEVICE; ASSIGN; RANK; **SEARCH** ; DOCUMENT;
CONTAIN; **WORD** ; EXTRACT; **WORD** ; CHARACTERISTIC; INFORMATION; CONTAIN;
PREFIX ; INDEPENDENT; **WORD**
Derwent Class: T01
International Patent Class (Main): G06F-017/30
File Segment: EPI
Manual Codes (EPI/S-X): T01-J05B3

14/9/5 (Item 5 from file: 350)
DIALOG(R)File 350:Derwent WPIX
(c) 2004 Thomson Derwent. All rts. reserv.

014117151 **Image available**
WPI Acc No: 2001-601363/200168
XRPX Acc No: N01-448588

Acoustic fast match processing method for speech recognition system, involves determining top ranking words based on computed path score and derived penalty score of each word in input text

Patent Assignee: INT BUSINESS MACHINES CORP (IBM)

Inventor: NOVAK M; PICHENY M

Number of Countries: 001 Number of Patents: 001

Patent Family:

| Patent No | Kind | Date | Applicat No | Kind | Date | Week |
|------------|------|----------|-------------|------|----------|----------|
| US 6275801 | B1 | 20010814 | US 98184870 | A | 19981103 | 200168 B |

Priority Applications (No Type Date): US 98184870 A 19981103

Patent Details:

| Patent No | Kind | Lan | Pg | Main IPC | Filing Notes |
|------------|------|-----|----|-------------|--------------|
| US 6275801 | B1 | | 10 | G10L-015/14 | |

Abstract (Basic): US 6275801 B1

NOVELTY - A penalty **score** for each **word** is derived based on the occurrence priority of each **word** in an input text, and a path **score** for each **word** is computed. The computed path score are combined with the derived penalty score to form a combined score. The combined score is tested against a threshold to determine top **ranking words**.

DETAILED DESCRIPTION - An INDEPENDENT CLAIM is also included for computer program device.

USE - For speech recognition system.

ADVANTAGE - A **word** is **searched** based on the computed path score and the derived penalty score depending on the occurrence priority of each **word** in an input text, thus improving the execution speed of an acoustic fast **match**, irrespective of the vocabulary size.

DESCRIPTION OF DRAWING(S) - The figure shows the partial construction of an asynchronous tree structure with associated penalty scores associated with each node.

pp; 10 DwgNo 3/4
Title Terms: ACOUSTIC; FAST; **MATCH** ; PROCESS; METHOD; SPEECH; RECOGNISE;
SYSTEM; DETERMINE; TOP; RANK; **WORD** ; BASED; COMPUTATION; PATH; SCORE;
DERIVATIVE ; PENALTY; SCORE; **WORD** ; INPUT; TEXT
Derwent Class: P86; W04

International Patent Class (Main): G10L-015/14
File Segment: EPI; EngPI
Manual Codes (EPI/S-X): W04-V01

14/9/6 (Item 6 from file: 350)
DIALOG(R)File 350:Derwent WPIX
(c) 2004 Thomson Derwent. All rts. reserv.

013905887
WPI Acc No: 2001-390100/200141
XRPX Acc No: N01-287001

Computer implemented method of searching using natural language
identifies topic, prefix and postfix words in a text string and
scores search results on the occurrence of these words

Patent Assignee: QJUNCTION TECHNOLOGY INC (QJUN-N); BASIR O (BASI-I);

KARRAY F (KARR-I); LEE V W L (LEE-V-I); SEMOTOK C (SEMO-I)

Inventor: BASIR O; KARRAY F; LEE V W L; SEMOTOK C; LEE V

Number of Countries: 091 Number of Patents: 003

Patent Family:

| Patent No | Kind | Date | Applicat No | Kind | Date | Week |
|----------------|------|----------|---------------|------|----------|----------|
| WO 200142981 | A2 | 20010614 | WO 2000IB2009 | A | 20001206 | 200141 B |
| AU 200122128 | A | 20010618 | AU 200122128 | A | 20001206 | 200161 |
| US 20010044720 | A1 | 20011122 | US 99169414 | A | 19991207 | 200176 |
| | | | US 2001732190 | A | 20010226 | |

Priority Applications (No Type Date): US 99169414 P 19991207; US 2001732190
A 20010226

Patent Details:

Patent No Kind Lan Pg Main IPC Filing Notes

WO 200142981 A2 E 43 G06F-017/30

Designated States (National): AE AL AM AT AU AZ BA BB BG BR BY CA CH CN
CR CU CZ DE DK DM EE ES FI GB GD GE GH GM HR HU ID IL IN IS JP KE KG KP
KR KZ LC LK LR LS LT LU LV MA MD MG MK MN MW MX NO NZ PL PT RO RU SD SE
SG SI SK SL TJ TM TR TT TZ UA UG US UZ VN YU ZA ZW

Designated States (Regional): AT BE CH CY DE DK EA ES FI FR GB GH GM GR
IE IT KE LS LU MC MW MZ NL OA PT SD SE SL SZ TR TZ UG ZW

AU 200122128 A G06F-017/30 Based on patent WO 200142981

US 20010044720 A1 G10L-015/04 Provisional application US 99169414

Abstract (Basic): WO 200142981 A2

NOVELTY - A search string is analyzed to identify topic words
and prefix and postfix descriptions. A database is searched for
the topic words and the prefix and postfix descriptions and
results are scored in dependence on the occurrence of the identified
topic words and the prefix and postfix descriptions. Additional
value is given to a search result with an exact match of the word
order found in the prefix description and the topic word (s).

DETAILED DESCRIPTION - An INDEPENDENT CLAIM is included for a
computer implemented system for searching a natural language string.

USE - Information retrieval .

ADVANTAGE - Streamlines and cost-effective way of searching using
natural English language inputs.

pp; 43 DwgNo 0/2

Title Terms: COMPUTER; IMPLEMENT; METHOD; SEARCH ; NATURAL; LANGUAGE;
IDENTIFY; TOPIC; PREFIX ; WORD ; TEXT; STRING; SCORE; SEARCH ; RESULT;
OCCUR; WORD

Derwent Class: P86; T01

International Patent Class (Main): G06F-017/30; G10L-015/04

File Segment: EPI; EngPI

Manual Codes (EPI/S-X): T01-J05B1; T01-J05B3; T01-J16C3
? t14/9/13,15-20

14/9/13 (Item 13 from file: 347)
DIALOG(R)File 347:JAPIO
(c) 2004 JPO & JAPIO. All rts. reserv.

07028255 **Image available**
SPEECH RECOGNITION DEVICE, SPEECH RECOGNITION METHOD AND RECORDING MEDIUM

PUB. NO.: 2001-255889 [JP 2001255889 A]
PUBLISHED: September 21, 2001 (20010921)
INVENTOR(s): HELMUT LUCKE
MINAMINO KATSUKI
ASANO KOJI
OGAWA HIROAKI
APPLICANT(s): SONY CORP
APPL. NO.: 2000-069698 [JP 200069698]
FILED: March 14, 2000 (20000314)
INTL CLASS: G10L-015/18; G06F-003/16

ABSTRACT

PROBLEM TO BE SOLVED: To prevent speech recognition precision from being degraded due to unknown **words** .

SOLUTION: A dictionary database 6 stores the **words** being objects of speech recognition and also a **word** dictionary in which **suffix words** being phonemes and phoneme trains constituting unknown **words** and classify the unknown **words** for every part of speech are registered. A **matching** section 4 connects acoustic models of an acoustic model database 5 based on the **word** dictionary and computes a score using a group of the featured values outputted by a feature extracting section 3 based on the connected acoustic models. Then, the section 4 selects a series of **words** being speech recognition **results** based on the **score** .

COPYRIGHT: (C)2001,JPO

14/9/15 (Item 15 from file: 347)
DIALOG(R)File 347:JAPIO
(c) 2004 JPO & JAPIO. All rts. reserv.

05420111 **Image available**
INFORMATION RETRIEVAL DEVICE

PUB. NO.: 09-034911 [JP 9034911 A]
PUBLISHED: February 07, 1997 (19970207)
INVENTOR(s): HIRAOKA NAOMI
ANDO MAKOTO
YAMASHITA AKIO
AIHARA KAZUO
KITA TATSUOMI
MATSUO HIROKO
KAWAMOTO SHINJI
YAMAGUCHI HIROSHI
APPLICANT(s): FUJI XEROX CO LTD [359761] (A Japanese Company or Corporation), JP (Japan)
APPL. NO.: 07-202747 [JP 95202747]
FILED: July 18, 1995 (19950718)
INTL CLASS: [6] G06F-017/30; G06F-017/21

JAPIO CLASS: 45.4 (INFORMATION PROCESSING -- Computer Applications)

ABSTRACT

PROBLEM TO BE SOLVED: To easily presume the contents of a **retrieved** document and to quickly select a **retrieval** document by displaying the number of times of the appearance of the **retrieval word** of **retrieved** information for the respective documents.

SOLUTION: A **retrieval** condition storage part 3 stores the **retrieval** conditions such as the **retrieval word**, **derivation** conditions, a **sorting** order and filtering conditions, etc., inputted from an input processing part 2 and a **retrieval** processing part 4 **retrieves** processing of the plural documents stored in a document storage part 1 corresponding to the **retrieval word** of the **retrieval** conditions stored in the storage part 3, the compound word of the **retrieval word** in a range derived by the **derivation** conditions and the relating word of the **retrieval word**. In this case, a **retrieval word** counting part 5 respectively counts the number of times of the appearance of the **retrieval word** appearing in the document by the **retrieval** processing, the compound word of the **retrieval word** in the range derived by the **derivation** conditions and the relating word of the **retrieval word**. A **retrieved** result storage part 6 stores the number of times of the appearance of the **retrieval word**, the compound word in the range derived by the **derivation** conditions and the relating word counted in the **retrieval word** counting part 5 along with the document name of the **retrieved** document as **retrieved** results.

14/9/16 (Item 16 from file: 347)

DIALOG(R)File 347:JAPIO

(c) 2004 JPO & JAPIO. All rts. reserv.

04044018 **Image available**

KANA/KANJI CONVERTER

PUB. NO.: 05-035718 [JP 5035718 A]

PUBLISHED: February 12, 1993 (19930212)

INVENTOR(s): TAKAHASHI FUMINO

APPLICANT(s): NEC CORP [000423] (A Japanese Company or Corporation), JP
(Japan)

APPL. NO.: 03-189635 [JP 91189635]

FILED: July 30, 1991 (19910730)

INTL CLASS: [5] G06F-015/20

JAPIO CLASS: 45.4 (INFORMATION PROCESSING -- Computer Applications)

JAPIO KEYWORD: R139 (INFORMATION PROCESSING -- Word Processors)

JOURNAL: Section: P, Section No. 1560, Vol. 17, No. 326, Pg. 86, June
21, 1993 (19930621)

ABSTRACT

PURPOSE: To improve the KANA (Japanese syllabary)/KANJI (Chinese character) converting efficiency by **sorting** the words connected with the **prefixes** and the **suffixes** for each **prefix** and **suffix** in the form of a table and also dividing the words connected to the **prefixes** and **suffixes** respectively into groups to register them into an example table.

CONSTITUTION: The input HIRAGANA (cursive form of Japanese syllabary) **character strings** are converted into a KANA-KANJI sentence through a converting part 3, a dictionary 4, a **prefix** table 51, and a **suffix** table 52. The dictionary 4 contains the words registered for KANA/KANJI conversion and gives the KANJI information or the pert-of-speech

information to the input **character strings** . Both tables 51 and 52 sort the **prefixes** and **suffixes** added to the **words** in accordance with the part-of-speech information of the **words** contained in the dictionary 4. Furthermore a **prefix** example table 61 or a **suffix** example table 62 is **retrieved** for the candidates consisting of the **prefixes** and the **words** or the **words** and the **suffixes** . These candidates are displayed as the conversion candidates.

14/9/17 (Item 17 from file: 347)
DIALOG(R)File 347:JAPIO
(c) 2004 JPO & JAPIO. All rts. reserv.

03175456 **Image available**
KANJI/KANJI CONVERTING DEVICE

PUB. NO.: 02-150956 [JP 2150956 A]
PUBLISHED: June 11, 1990 (19900611)
INVENTOR(s): KOYAMA YASUO
APPLICANT(s): SEIKO EPSON CORP [000236] (A Japanese Company or Corporation)
, JP (Japan)
APPL. NO.: 63-304655 [JP 88304655]
FILED: December 01, 1988 (19881201)
INTL CLASS: [5] G06F-015/20
JAPIO CLASS: 45.4 (INFORMATION PROCESSING -- Computer Applications)
JAPIO KEYWORD:R139 (INFORMATION PROCESSING -- **Word** Processors)
JOURNAL: Section: P, Section No. 1097, Vol. 14, No. 398, Pg. 157,
August 28, 1990 (19900828)

ABSTRACT

PURPOSE: To realize the more natural KANA (Japanese syllabary)/KANJI (Chinese character) conversion and to improve the **hit rate** by noting only the independence between two paragraphs to learn the co-occurrence of paragraphs and ensuring the learning of paragraphs which can withstand even the change of the **affixal words** .

CONSTITUTION: A KANA/KANJI conversion part 8 **retrieves** an independent **word** dictionary 12 and an **affixal** work dictionary 11 based on the input KANA **character string** and at the same time stores the **word** candidates into a **word** candidates storing part 10. Then the part 8 checks whether an independent **word** is connected to an **affixal word** or not by reference to a **word** connection table 14. Thus an independent- **affixal** work group is formed to include all **affixal words** which are connected to the independent **words** and the **affixal words** which can be connected to the former **affixal words** . These **words** are composed into the paragraphs and stored in a paragraph candidate storing part 9. Then the part 8 recognizes the punctuation of the connectable paragraph strings where the paragraph of the least total cost is defined as the final paragraph as the result of the KANA/KANJI conversion.

14/9/18 (Item 18 from file: 347)
DIALOG(R)File 347:JAPIO
(c) 2004 JPO & JAPIO. All rts. reserv.

02649675
JAPANESE SYLLABARY TO CHINESE CHARACTER CONVERSION SYSTEM

PUB. NO.: 63-266575 [JP 63266575 A]
PUBLISHED: November 02, 1988 (19881102)

INVENTOR(s): ITO TOMIHIRO
APPLICANT(s): ALPS ELECTRIC CO LTD [001009] (A Japanese Company or Corporation), JP (Japan)
APPL. NO.: 62-100910 [JP 87100910]
FILED: April 23, 1987 (19870423)
INTL CLASS: [4] G06F-015/20
JAPIO CLASS: 45.4 (INFORMATION PROCESSING -- Computer Applications)
JAPIO KEYWORD: R139 (INFORMATION PROCESSING -- Word Processors)
JOURNAL: Section: P, Section No. 834, Vol. 13, No. 85, Pg. 53, February 27, 1989 (19890227)

ABSTRACT

PURPOSE: To reduce the display of strange 'KANJI' (Chinese character) to which an unnecessary **suffix** is added by handling **suffixes** similarly to independent **words**, and comparing the **suffixes** with other independent **words** in consideration of **word** length and selecting 'KANJI'.

CONSTITUTION: The **suffixes** are considered to be equal to independent **words** and when the independent **word** of a subsequent paragraph is processed, a **suffix** is **retrieved** according to data on the last paragraph which is already determined and compared with independent homonymous **words**, thereby giving priority to the **suffix** at the time of similar paragraph length. The, the **suffix** is handled as a **sort** of independent **word** and **searched** for as an independent **word** according to the relation with the last paragraph which is already determined to perform conversion in consideration of the **word** length, so the display of a paragraph of strange 'KANJI' to which a **suffix** is added unnecessarily is reduced, the correct **answer rate** of the conversion to 'KANJI' is increased, and the efficiency of document creation is improved.

14/9/19 (Item 19 from file: 347)

DIALOG(R) File 347:JAPIO

(c) 2004 JPO & JAPIO. All rts. reserv.

01162428 **Image available**

"KANA" (JAPANESE SYLLABARY) - CHINESE CHARACTER CONVERTING PROCESSOR

PUB. NO.: 58-099828 [JP 58099828 A]
PUBLISHED: June 14, 1983 (19830614)
INVENTOR(s): YAMAUCHI YOSHITOSHI
HAYASHI HIROKAWA
APPLICANT(s): RICOH CO LTD [000674] (A Japanese Company or Corporation), JP (Japan)
APPL. NO.: 56-197383 [JP 81197383]
FILED: December 08, 1981 (19811208)
INTL CLASS: [3] G06F-003/02; G06F-015/38
JAPIO CLASS: 45.3 (INFORMATION PROCESSING -- Input Output Units); 29.4 (PRECISION INSTRUMENTS -- Business Machines); 30.2 (MISCELLANEOUS GOODS -- Sports & Recreation); 45.4 (INFORMATION PROCESSING -- Computer Applications)
JAPIO KEYWORD: R106 (INFORMATION PROCESSING -- Kanji Information Processing)
JOURNAL: Section: P, Section No. 221, Vol. 07, No. 202, Pg. 82, September 07, 1983 (19830907)

ABSTRACT

PURPOSE: To simplify an operation of a device, by constituting a storage part of **output ranking** information, of plural parts which are capable of storing different information in each field, designating its plural parts, changing the **output ranking** of a homonym, and reducing the

number of times of operation of the homonym.

CONSTITUTION: From a keyboard device 11, a Japanese sentence is stored as a 'Kana' (Japanese syllabary) character storing temporarily in a 'Kana' register 12, and a general independent word dictionary 18A, a general prefix dictionary 19A, a general suffix dictionary 20A, etc., which have been divided by each field are retrieved by control of a controlling circuit 13. By this retrieval, contents of the register 12 are converted to a sentence containing a Chinese character and are displayed on a display 15. On these respective dictionaries 18A, 19A and 20A, an independent word buffer 26, a prefix buffer 28 and a suffix buffer 30 are provided, and by each ranking controller 27, 29 and 31, the output ranking is changed as to the dictionary 18A, a proper noun dictionary 18B, a numeral dictionary 18C, the dictionaries 19A, 20A, proper noun prefix and suffix dictionaries 19B, 20B, and preposition and post-position auxiliary numeral tables 19C, 20C, and the number of times of operation of a homonym is reduced.

14/9/20 (Item 20 from file: 347)

DIALOG(R)File 347:JAPIO

(c) 2004 JPO & JAPIO. All rts. reserv.

01113981 **Image available**

KANA (JAPANESE SYLLABARY)-KANJI (CHINESE CHARACTER) CONVERSION PROCESSING DEVICE

PUB. NO.: 58-051381 [JP 58051381 A]

PUBLISHED: March 26, 1983 (19830326)

INVENTOR(s): YAMAUCHI YOSHITOSHI

HAYASHI HIROKAWA

APPLICANT(s): RICOH CO. LTD [000674] (A Japanese Company or Corporation), JP (Japan)

APPL. NO.: 56-149737 [JP 81149737]

FILED: September 22, 1981 (19810922)

INTL CLASS: [3] G06F-015/38; G06F-003/02

JAPIO CLASS: 45.4 (INFORMATION PROCESSING -- Computer Applications); 30.2 (MISCELLANEOUS GOODS -- Sports & Recreation); 45.3 (INFORMATION PROCESSING -- Input Output Units)

JAPIO KEYWORD: R106 (INFORMATION PROCESSING -- Kanji Information Processing)

JOURNAL: Section: P, Section No. 203, Vol. 07, No. 134, Pg. 163, June 11, 1983 (19830611)

ABSTRACT

PURPOSE: To determine an output order matched to the use inclination, by exchanging output order information between a word, which is selectively outputted from the output, and a homophone of the word which is outputted just before this output.

CONSTITUTION: An independent word dictionary is divided into three parts, namely, a general independent word dictionary 18A, a proper noun dictionary 18B, and a numeral dictionary 18C; and in accordance with this division, a prefix table is divided into three parts, namely, a general prefix table 19A, a proper noun prefix dictionary 19B, and a prepositive auxiliary numeral table 19C, and a suffix table is divided into three parts, namely, a general suffix table 20A, a proper noun suffix table 20B, and a postpositive auxiliary numeral table 20C. These fractionized dictionaries and tables are so constituted that corresponding dictionaries and tables can be selected to be used for collation by interlocking changeover switches S(sub 1), S(sub 2), and S(sub 3). Thus,

clauses are classified into categories and are inputted together with pertinent category information to improve the **answer rate** (correct conversion **rate**) of analysis.
?

2000

File 9:Business & Industry(R) Jul/1994-2004/Sep 20
(c) 2004 The Gale Group
File 16:Gale Group PROMT(R) 1990-2004/Sep 21
(c) 2004 The Gale Group
File 47:Gale Group Magazine DB(TM) 1959-2004/Sep 21
(c) 2004 The Gale group
File 88:Gale Group Business A.R.T.S. 1976-2004/Sep 20
(c) 2004 The Gale Group
File 148:Gale Group Trade & Industry DB 1976-2004/Sep 21
(c)2004 The Gale Group
File 160:Gale Group PROMT(R) 1972-1989
(c) 1999 The Gale Group
File 275:Gale Group Computer DB(TM) 1983-2004/Sep 21
(c) 2004 The Gale Group
File 570:Gale Group MARS(R) 1984-2004/Sep 21
(c) 2004 The Gale Group
File 621:Gale Group New Prod.Annou.(R) 1985-2004/Sep 21
(c) 2004 The Gale Group
File 636:Gale Group Newsletter DB(TM) 1987-2004/Sep 21
(c) 2004 The Gale Group
File 649:Gale Group Newswire ASAP(TM) 2004/Sep 15
(c) 2004 The Gale Group
File 1:ERIC 1966-2004/Jul 21
(c) format only 2004 The Dialog Corporation

| Set | Items | Description |
|-----|---------|--|
| S1 | 2706005 | QUERY? OR QUERIE? ? OR SUBQUER? OR SEARCH? OR FETCH? OR RETRIEV? OR TEXTSEARCH? |
| S2 | 1164755 | MATCH??? ? |
| S3 | 7160798 | WORD? ? OR TERM? ? |
| S4 | 7351 | CHARACTERSTRING? OR CHARACTER? ?(2N)STRING? ? OR WORDSTEM? OR MORPHEME? OR WORDELEMENT? OR BASETERM? OR BASEWORD? OR LEXEME? |
| S5 | 15402 | S3(2N) (STEM? ? OR ELEMENT? ? OR BASE OR BASES) |
| S6 | 351361 | SUFFIX? OR PREFIX? OR DERIVATI? OR AFFIX? OR POSTFIX? OR TRUNCAT? OR LEFTTRUNCAT? OR RIGHTTRUNCAT? |
| S7 | 9466344 | RANK? OR RATE OR RATES OR RATED OR RATING? OR SORT??? ? OR SCOR??? ? OR VALUATION? OR TALLY? OR TALLIE? ? OR WEIGH? |
| S8 | 337809 | S7(3N) (RESULT? OR HITLIST? OR REFERENCE? OR RETRIEV? OR HIT OR HITS OR OUTPUT? OR OUT()PUT? ? OR RESPONSE? ? OR ANSWER? ? OR REPLIE? ? OR REPLY?) |
| S9 | 249356 | S2:S4(3N)S7 |
| S10 | 69704 | S1:S2(5N)S3:S5 |
| S11 | 1455 | S10(S)S6 |
| S12 | 103 | S11(S)S8:S9 |
| S13 | 13 | S12/2001:2004 |
| S14 | 90 | S12 NOT S13 |
| S15 | 54 | RD (unique items) |

15/3,K/1 (Item 1 from file: 16)

DIALOG(R)File 16:Gale Group PROMT(R)

(c) 2004 The Gale Group. All rts. reserv.

08030957 Supplier Number: 66101423 (USE FORMAT 7 FOR FULLTEXT)
Inktomi Serves Up Smart Searches -- Looking for excellent retrieval
quality, power and flexibility in a Web site search engine? Inktomi
Search Software has what you've been searching for. (Software
Review) (Evaluation)

Rappoport, Avi
Network Computing, p54
Oct 16, 2000

Language: English Record Type: Fulltext
Article Type: Evaluation
Document Type: Magazine/Journal; Trade
Word Count: 4156

... fields for searching in the text, title and date ranges, and options for showing the **results** or **sorting** by date. However, the advanced-search form does not make a multiterm search as simple...

...query. It does a good job of locating the singular or plural form of a **search term** in English, and allows both left and right **truncation**, so a search for *workcom* would yield networkcomputing as a **result**.

To **sort** the **results**, Searchbutton emphasizes pages with the search terms in the title and those with several instances...

15/3,K/6 (Item 6 from file: 16)
DIALOG(R)File 16:Gale Group PROMT(R)
(c) 2004 The Gale Group. All rts. reserv.

06477212 Supplier Number: 55109370 (USE FORMAT 7 FOR FULLTEXT)
Pagis Pro 3.0 Scanning Suite. (from ScanSoft) (Software Review) (Evaluation)
Haskin, David
PC Magazine, p66
August 1, 1999
Language: English Record Type: Fulltext
Article Type: Evaluation
Document Type: Magazine/Journal; General Trade
Word Count: 304

... has indexed are slightly stronger than PageKeeper's. Like its competitor, Pagis provides a relevancy- **ranked** list of search **results** and supports fuzzy searches. In addition, the software can **search** for similar sounding **words** and **derivations** of words, such as the same verb in different tenses.

This program is full of...

15/3,K/7 (Item 1 from file: 47)
DIALOG(R)File 47:Gale Group Magazine DB(TM)
(c) 2004 The Gale group. All rts. reserv.

05489447 SUPPLIER NUMBER: 54796472 (USE FORMAT 7 OR 9 FOR FULL TEXT)
AVAILABILITY AND COST OF WEB-BASED BIBLIOGRAPHIC SEARCH SERVICES. (World Wide Web)
Library Technology Reports, 35, 1, 7
Jan, 1999
ISSN: 0024-2586 LANGUAGE: English RECORD TYPE: Fulltext; Abstract
WORD COUNT: 45904 LINE COUNT: 04199

... databases are automatically selected and searched. Multiple databases are searched simultaneously and duplicate records removed.

Search terms are typed into dialog boxes that correspond to fields to be searched. Thus, search options for the industry news category include company name, **words** in title, and entire text. **Searches** can be limited to full-text articles and to specified date ranges. Dialog Select automatically **searches** related **terms**. Right truncation of **search terms** is permitted. **Retrieved** records are **sorted** by date and initially displayed in a brief title format that includes the publication date...for periodic re-execution. Brief records can be selected to display more

complete information. Search **results** can be **sorted** by author, date, or other parameters. Displayed records can be marked for printing or downloading...will retrieve words of similar spelling; word expansion, which Suggests specific meanings and spellings for **search terms** ; **term weighting** ; and a natural language option, which will expand a **search** to include **words** with related meanings.

MD Consult (www.mdconsult.com) is a web-based information service designed...among other features. Searches can be limited by publication date, document type, or language. Search **results** can be **sorted** by date or ranked for relevance.

Knowledge Web (www.knowledgeweb.com) is a medical information...among other features. Searches can be limited by publication date, document type, or language. Search **results** can be **sorted** by date or ranked for relevance.

Community of Science (www.cos.com) was formed by...

15/3,K/9 (Item 3 from file: 47)

DIALOG(R)File 47:Gale Group Magazine DB(TM)
(c) 2004 The Gale group. All rts. reserv.

05391070 SUPPLIER NUMBER: 54864176 (USE FORMAT 7 OR 9 FOR FULL TEXT)
The Federal Register Free on GPO Access. (Government Printing Office)
Gordon-Murnane, Laura
Searcher, 7, 6, 46
June, 1999
ISSN: 1070-4795 LANGUAGE: English RECORD TYPE: Fulltext
WORD COUNT: 6108 LINE COUNT: 00621

... Truncation: The asterisk (*) acts as the truncation operator. It can be used to truncate a **word** and search for all **words** within a **word stem**.

6. Relevance **Ranking** : After completing a search, the WAIS server automatically displays a relevancy ranking for each retrieved...

15/3,K/10 (Item 4 from file: 47)

DIALOG(R)File 47:Gale Group Magazine DB(TM)
(c) 2004 The Gale group. All rts. reserv.

05354745 SUPPLIER NUMBER: 54467218 (USE FORMAT 7 OR 9 FOR FULL TEXT)
Facts On File World News CD-ROM. (electronic reference) (Software Review) (Evaluation) (Brief Article)
Roberts, Randall L.
Reference & User Services Quarterly, 38, 1, 84(1)
Fall, 1998
DOCUMENT TYPE: Evaluation Brief Article ISSN: 1094-9054
LANGUAGE: English RECORD TYPE: Fulltext
WORD COUNT: 906 LINE COUNT: 00078

... full-text search features such as Boolean and proximity searching, exact phrase searching, and wildcard (**truncation**) searching are available. PLWeb-CD also offers stemming (finding word variants), concept operators (finding statistically related **terms**), fuzzy **searching** (finding similar spellings), intelligent searching (search engine "interprets" a user's natural language query), and even relevancy **ranking** of **retrieved** documents. The skilled searcher can execute most of these features in a single command line...

15/3,K/11 (Item 5 from file: 47)
DIALOG(R)File 47:Gale Group Magazine DB(TM)
(c) 2004 The Gale group. All rts. reserv.

05156573 SUPPLIER NUMBER: 19539822 (USE FORMAT 7 OR 9 FOR FULL TEXT)
Client/server products, additional evaluations. (The graphical user interface (GUI) in library products, Part 2)
Matthews, Joseph R.
Library Technology Reports, v33, n1, p43(52)
Jan-Feb, 1997
ISSN: 0024-2586 LANGUAGE: English RECORD TYPE: Fulltext; Abstract
WORD COUNT: 12139 LINE COUNT: 01216

... of the OPAC. The Windows version of the OPAC provides automatic truncation, spell checking and **weighting** of keyword search **results** using relevancy **ranking**.

To conduct a subject search, the user first selects the type of search to be...

15/3,K/13 (Item 7 from file: 47)
DIALOG(R)File 47:Gale Group Magazine DB(TM)
(c) 2004 The Gale group. All rts. reserv.

05079151 SUPPLIER NUMBER: 19581031 (USE FORMAT 7 OR 9 FOR FULL TEXT)
Comstow Information Services. (Vendors of Integrated Library Systems for Minicomputers and Mainframes: An Industry Report, part 1)
Saffady, William
Library Technology Reports, v33, n2, p185(8)
March-April, 1997
ISSN: 0024-2586 LANGUAGE: English RECORD TYPE: Fulltext; Abstract
WORD COUNT: 3760 LINE COUNT: 00336

... responds to multi-field searches with a count of retrieved items plus a display of **search terms** in their surrounding context.

The **searching** module supports complex retrieval operations, including relational expressions, Boolean operators (AND, OR, NOT, and XOR), nested search terms, wildcard **characters** in search strings, and **sorting** instructions for **retrieved** records. Search commands can be saved for repeated execution. Searchers can examine a thesaurus for...

15/3,K/14 (Item 8 from file: 47)
DIALOG(R)File 47:Gale Group Magazine DB(TM)
(c) 2004 The Gale group. All rts. reserv.

05076980 SUPPLIER NUMBER: 19581032 (USE FORMAT 7 OR 9 FOR FULL TEXT)
Ex Libris Limited. (Vendors of Integrated Library Systems for Minicomputers and Mainframes: An Industry Report, part 1)
Saffady, William
Library Technology Reports, v33, n2, p193(11)
March-April, 1997
ISSN: 0024-2586 LANGUAGE: English RECORD TYPE: Fulltext; Abstract
WORD COUNT: 5303 LINE COUNT: 00465

... holdings information. The command mode supports complex retrieval capabilities, including command stacking, right and left **truncation** of **search terms**, Boolean operators, relational expressions, proximity operators, and **ranking** of search **results** by frequency of occurrence of **search terms**. **Searches** can be limited by publication date, language,

or other library-defined parameters.
Search results can...

15/3,K/15 (Item 9 from file: 47)
DIALOG(R)File 47:Gale Group Magazine DB(TM)
(c) 2004 The Gale group. All rts. reserv.

04833456 SUPPLIER NUMBER: 19761239 (USE FORMAT 7 OR 9 FOR FULL TEXT)
Vendors of integrated library systems for minicomputers and mainframes: an industry report, part 2. (part 1: Contec Data Systems, Data Research Associates, Endeavor Information Systems, EOS International, Fretwell Downing Informatics) (Company Profile)
Saffady, William
Library Technology Reports, v33, n3, p277(50)
May-June, 1997
DOCUMENT TYPE: Company Profile ISSN: 0024-2586 LANGUAGE: English
RECORD TYPE: Fulltext; Abstract
WORD COUNT: 22345 LINE COUNT: 01943

... desired, the operator can use specified characters to mark certain keywords as essential to a **search** or more important than other **words** .

Term truncation permits root- **word searches** .

For keyword **searches** , Voyager **ranks retrieved** records by their presumed relevance. Relevance ranking, which resembles methods employed by Internet search engines...also provides such unusual features as phrase searching, the ability to differentiate essential and important **search terms** , and relevance **ranking of retrieved** records, a feature that is supported by other new integrated systems discussed in this issue...users, while experienced searchers and library staff members have access to Boolean operators, relational expressions, **term truncation** , hypertext **searching** , and other advanced retrieval features. Drawing on a powerful search engine developed by Excalibur Technologies, the Q Series is particularly notable for relevance **ranking of retrieved** records, soundex searching, automatic substitution of synonyms and variant spellings, and for accepting search statements...

15/3,K/18 (Item 12 from file: 47)
DIALOG(R)File 47:Gale Group Magazine DB(TM)
(c) 2004 The Gale group. All rts. reserv.

04641376 SUPPLIER NUMBER: 18848837 (USE FORMAT 7 OR 9 FOR FULL TEXT)
Ovid Web gateway: nobody does it better.
Jacso, Peter
Online, v20, n6, p24(7)
Nov-Dec, 1996
ISSN: 0146-5422 LANGUAGE: English RECORD TYPE: Fulltext; Abstract
WORD COUNT: 2947 LINE COUNT: 00230

... derivative mapping. Ovid automatically creates and conducts a search based on the user's query, **retrieves** the qualifying records, **ranks** the major subject headings in some of those records, and presents the user with up to ten of the most often occurring subject headings or the permuted index of a **term** . The original user **query** is also displayed, allowing the user to do her own keyword search.

Ovid does not...
? t15/3,k/23-24,29,36

15/3,K/23 (Item 17 from file: 47)

DIALOG(R)File 47:Gale Group Magazine DB(TM)
(c) 2004 The Gale group. All rts. reserv.

03970501 SUPPLIER NUMBER: 14089265 (USE FORMAT 7 OR 9 FOR FULL TEXT)
Dialog's RANK command: building and mining the data mountain.
Basch, Reva
Online, v17, n4, p28(8)
July, 1993
ISSN: 0146-5422 LANGUAGE: ENGLISH RECORD TYPE: FULLTEXT; ABSTRACT
WORD COUNT: 3785 LINE COUNT: 00285

... in the top three-ranked counties.

Possible Confusion Alert (PCA): RANK defaults to the current **answer** set. To **RANK** on an earlier set, you must precede the set number with an "S," e.g...

...It will display the distinctly unhelpful message, Rank fields found in 1 records -- 1 unique **terms**. **Searchers** would be much better served with an error message, something like "Precede set number with..."

...specify the set number, even for the default set, but doesn't require the S- **prefix** --RANK doesn't require the set number, but does require an "S" if you do...

15/3,K/24 (Item 18 from file: 47)
DIALOG(R)File 47:Gale Group Magazine DB(TM)
(c) 2004 The Gale group. All rts. reserv.

03900654 SUPPLIER NUMBER: 14258505 (USE FORMAT 7 OR 9 FOR FULL TEXT)
INMAGIC Plus for Libraries. (Microcomputer-Based Automated Library Systems: New Series, Part 2, 1993) (Software Review) (Evaluation)
Library Technology Reports, v29, n3, p327(8)
May-June, 1993
DOCUMENT TYPE: Evaluation ISSN: 0024-2586 LANGUAGE: ENGLISH
RECORD TYPE: FULLTEXT; ABSTRACT
WORD COUNT: 1912 LINE COUNT: 00171

... of the record. (See figure 3.)

The user can specify boolean AND, OR, and NOT **search terms**, but must enter the characters &, /, &-" for the boolean AND, OR, and NOT, respectively. **Truncation** and proximity searching are available when a search is entered or later modified. Searches can...

...range. Left to right or character by character phrase and keyword searching is supported. Search **results** can be **sorted**, saved, and recalled for later use.

A command option allows the user to enter searches...

15/3,K/29 (Item 1 from file: 148)
DIALOG(R)File 148:Gale Group Trade & Industry DB
(c)2004 The Gale Group. All rts. reserv.

12817764 SUPPLIER NUMBER: 67316055 (USE FORMAT 7 OR 9 FOR FULL TEXT)
Gale Goes Free.
O'Leary, Mick
EContent, 23, 6, 71
Dec, 2000
ISSN: 1525-2531 LANGUAGE: English RECORD TYPE: Fulltext

WORD COUNT: 1616 LINE COUNT: 00135

... in a phrase entered without quotation marks are searched with an implied Boolean OR. Search **results** are **sorted** by relevance ranking, but there is no date sorting option. Format is text, not image...

15/3,K/36 (Item 8 from file: 148)
DIALOG(R)File 148:Gale Group Trade & Industry DB
(c)2004 The Gale Group. All rts. reserv.

07576484 SUPPLIER NUMBER: 15875813 (USE FORMAT 7 OR 9 FOR FULL TEXT)
Searching natural language systems: searchers know thy engine. (includes related article on new natural language search engines)
Feldman, Susan E.
Searcher, v2, n8, p34(5)
Oct, 1994
ISSN: 1070-4795 LANGUAGE: ENGLISH RECORD TYPE: FULLTEXT
WORD COUNT: 4390 LINE COUNT: 00349

... 4. Stemming: words can be automatically truncated or expanded to allow for plural/singular problems. **Word stems** are isolated and **matched** against **words** with the same stem.

5. Very frequent terms may be ignored entirely. WAIS ignores any...
? t15/3,k/38,40,44

15/3,K/38 (Item 10 from file: 148)
DIALOG(R)File 148:Gale Group Trade & Industry DB
(c)2004 The Gale Group. All rts. reserv.

07238930 SUPPLIER NUMBER: 14974306 (USE FORMAT 7 OR 9 FOR FULL TEXT)
PHARMSEARCH enhanced with images, additional indexing data.
Information Today, v11, n3, p 12(1)
March, 1994
ISSN: 8755-6286 LANGUAGE: ENGLISH RECORD TYPE: FULLTEXT
WORD COUNT: 450 LINE COUNT: 00042

... include: implied adjacency of terms, left-hand truncation, stringserach of text, expanded statistical analysis capability, **ranking** of search **results**, and expanded cross-file searching possibilities.
Additional Bibliographic Data
Five additional kinds of bibliographic data...

15/3,K/40 (Item 12 from file: 148)
DIALOG(R)File 148:Gale Group Trade & Industry DB
(c)2004 The Gale Group. All rts. reserv.

07202543 SUPPLIER NUMBER: 15021905 (USE FORMAT 7 OR 9 FOR FULL TEXT)
Testing database quality. (online bibliographic databases)
Cahn, Pamela
Database, v17, n1, p23(7)
Feb, 1994
ISSN: 0162-4105 LANGUAGE: ENGLISH RECORD TYPE: FULLTEXT; ABSTRACT
WORD COUNT: 3036 LINE COUNT: 00241

... shows the error rates for the modified list of truncated terms.
Note the shift in **ranking** as a **result** of **searching** for variations on **word** form.

None of the previously mentioned tables relate size of the database

to error rates...

15/3,K/44 (Item 16 from file: 148)
DIALOG(R)File 148:Gale Group Trade & Industry DB
(c)2004 The Gale Group. All rts. reserv.

06405828 SUPPLIER NUMBER: 13438889 (USE FORMAT 7 OR 9 FOR FULL TEXT)
New version of QUESTEL Plus software available. (Product Announcement)
Information Today, v10, n1, p1(2)
Jan, 1993
DOCUMENT TYPE: Product Announcement ISSN: 8755-6286 LANGUAGE:
ENGLISH RECORD TYPE: FULLTEXT
WORD COUNT: 591 LINE COUNT: 00047

... search for strings or terms in a very precise order and which are not otherwise **retrievable** .

Questel's new .. **RANK** (..RK) command lets users display results in descending order of occurrence of the search terms...
? t15/3,k/47,49-50

15/3,K/47 (Item 19 from file: 148)
DIALOG(R)File 148:Gale Group Trade & Industry DB
(c)2004 The Gale Group. All rts. reserv.

05072730 SUPPLIER NUMBER: 09825331
How effective is suffixing?
Harman, Donna
Journal of the American Society for Information Science, v42, n1, p7(9)
Jan, 1991
DOCUMENT TYPE: Evaluation ISSN: 0002-8231 LANGUAGE: ENGLISH
RECORD TYPE: ABSTRACT

...ABSTRACT: for any of the algorithms. A failure analysis suggested three modifications to ranking techniques: variable **weighting** of **term** variants, selective stemming depending on **query** length, and selective stemming depending on term importance. None of these modifications improved performance. Recommendations are made regarding the uses of **suffixing** in an online environment. (Reprinted by permission of the publisher.)

15/3,K/49 (Item 1 from file: 160)
DIALOG(R)File 160:Gale Group PROMT(R)
(c) 1999 The Gale Group. All rts. reserv.

00621164
Cuadra Assoc demonstrated a new online microcomputer data entry and retrieval system at the ASIS meeting 10/80.
Online January, 1981 p. 10,72

... or infrequent, users. STAR features are full Boolean logic, use of nested parentheses for complex **queries**, **truncated term searching**, use of controlled vocabulary, as well as free text searching, index displays, flexible online and offline print formats, and **sorting of output** .

15/3,K/50 (Item 1 from file: 275)
DIALOG(R)File 275:Gale Group Computer DB(TM)
(c) 2004 The Gale Group. All rts. reserv.

01496740 SUPPLIER NUMBER: 11875457 (USE FORMAT 7 OR 9 FOR FULL TEXT)
Living in parallel. (commercial uses of parallel systems)
Keyes, Jessica
AI Expert, v7, n2, p42(6)
Feb, 1992
ISSN: 0888-3785 LANGUAGE: ENGLISH RECORD TYPE: FULLTEXT; ABSTRACT
WORD COUNT: 3377 LINE COUNT: 00270

... of parallel processors for text retrieval is DowQuest's use of what is known as **term weights** . At the end of the first pass through the database, the user might choose two...

...again. The second search uses the entirety of the two articles (for example, 1,200 **searchable words**) as a thesaurus. It **ranks** those **words** according to **weight** and then gives the rarest **words** the most **weight** . The computer then **truncates** the list to the top 100 **words** , reinitiating the **search** . The net result is that users can find all applicable articles of interest in a...
?

File 696:DIALOG Telecom. Newsletters 1995-2004/Sep 20
(c) 2004 The Dialog Corp.
File 15:ABI/Inform(R) 1971-2004/Sep 21
(c) 2004 ProQuest Info&Learning
File 98:General Sci Abs/Full-Text 1984-2004/Jul
(c) 2004 The HW Wilson Co.
File 112:UBM Industry News 1998-2004/Jan 27
(c) 2004 United Business Media
File 141:Readers Guide 1983-2004/Jul
(c) 2004 The HW Wilson Co
File 484:Periodical Abs Plustext 1986-2004/Sep W2
(c) 2004 ProQuest
File 608:KR/T Bus.News. 1992-2004/Sep 21
(c)2004 Knight Ridder/Tribune Bus News
File 813:PR Newswire 1987-1999/Apr 30
(c) 1999 PR Newswire Association Inc
File 613:PR Newswire 1999-2004/Sep 21
(c) 2004 PR Newswire Association Inc
File 635:Business Dateline(R) 1985-2004/Sep 21
(c) 2004 ProQuest Info&Learning
File 810:Business Wire 1986-1999/Feb 28
(c) 1999 Business Wire
File 610:Business Wire 1999-2004/Sep 21
(c) 2004 Business Wire.
File 369:New Scientist 1994-2004/Sep W2
(c) 2004 Reed Business Information Ltd.
File 370:Science 1996-1999/Jul W3
(c) 1999 AAAS
File 20:Dialog Global Reporter 1997-2004/Sep 21
(c) 2004 The Dialog Corp.
File 624:McGraw-Hill Publications 1985-2004/Sep 20
(c) 2004 McGraw-Hill Co. Inc
File 634:San Jose Mercury Jun 1985-2004/Sep 19
(c) 2004 San Jose Mercury News
File 647:CMP Computer Fulltext 1988-2004/Sep W2
(c) 2004 CMP Media, LLC
File 674:Computer News Fulltext 1989-2004/Aug W4
(c) 2004 IDG Communications

| Set | Items | Description |
|-----|----------|--|
| S1 | 2849541 | QUERY? OR QUERIE? ? OR SUBQUER? OR SEARCH? OR FETCH? OR RETRIEV? OR TEXTSEARCH? |
| S2 | 2173581 | MATCH??? ? |
| S3 | 7921058 | WORD? ? OR TERM? ? |
| S4 | 2292 | CHARACTERSTRING? OR CHARACTER? ?(2N)STRING? ? OR WORDSTEM? OR MORPHEME? OR WORDELEMENT? OR BASETERM? OR BASEWORD? OR LEX-EME? |
| S5 | 13363 | S3(2N) (STEM? ? OR ELEMENT? ? OR BASE OR BASES) |
| S6 | 346027 | SUFFIX? OR PREFIX? OR DERIVATI? OR AFFIX? OR POSTFIX? OR TRUNCAT? OR LEFTTRUNCAT? OR RIGHTTRUNCAT? |
| S7 | 10495196 | RANK? OR RATE OR RATES OR RATED OR RATING? OR SORT??? ? OR SCOR??? ? OR VALUATION? OR TALLY? OR TALLIE? ? OR WEIGH? |
| S8 | 312507 | S7(3N) (RESULT? OR HITLIST? OR REFERENCE? OR RETRIEV? OR HIT OR HITS OR OUTPUT? OR OUT()PUT? ? OR RESPONSE? ? OR ANSWER? ? OR REPLIE? ? OR REPLY?) |
| S9 | 318368 | S2:S4(3N)S7 |
| S10 | 55702 | S1:S2(5N)S3:S5 |
| S11 | 737 | S10(S)S6 |
| S12 | 85 | S11(S)S8:S9 |
| S13 | 30 | S12/2001:2004 |

S14 55 S12 NOT S13
S15 42 RD (unique items)

15/3,K/12 (Item 12 from file: 15)
DIALOG(R)File 15:ABI/Inform(R)
(c) 2004 ProQuest Info&Learning. All rts. reserv.

01343416 99-92812
Ovid Web Gateway: Nobody does it better
Jacso, Peter
Online v20n6 PP: 24-31 Nov/Dec 1996
ISSN: 0146-5422 JRNL CODE: ONL
WORD COUNT: 2719

...TEXT: derivative mapping. Ovid automatically creates and conducts a search based on the user's query, **retrieves** the qualifying records, **ranks** the major subject headings in some of those records, and presents the user with up to ten of the most often occurring subject headings or the permuted index of a term. The original user **query** is also displayed, allowing the user to do her own keyword search.

Ovid does not...

15/3,K/15 (Item 15 from file: 15)
DIALOG(R)File 15:ABI/Inform(R)
(c) 2004 ProQuest Info&Learning. All rts. reserv.

01160558 98-09953
Robot-generated databases on the World Wide Web
Kimmel, Stacey
Database v19n1 PP: 40-49 Feb/Mar 1996
ISSN: 0162-4105 JRNL CODE: DTB
WORD COUNT: 4074

...TEXT: document..

Lycos offers two search forms: a simple search form, with a place to enter **search terms**, and a detailed form with additional options (Figure 2). (Figure 2 omitted) The detailed form lets users specify maximum hits (10, 20, 30, or 40 results per page). Options to **Match Any (OR) terms**, **All (AND) terms**, or a specified number of terms (from two to seven) are also provided, and matches...

... title only), standard (title, outline, abstract, URL), or verbose (ranking, title, number of links, outline, **matched words**, abstract, description, URL, file size, date updated). Hyphens and other non-alphanumeric characters are not...

...The Boolean connector NOT can be approximated by placing a hyphen (-) in front of a **search word**. This decreases the likelihood that the **word** will appear in the **search** results but does not eliminate it entirely. Although word adjacency cannot be specified in the **search** form, Lycos uses **word** proximity to **rank** documents. When **search words** appear close together in the text, the document receives a higher relevance **score** than when **words** appear further apart. Lycos employs automatic **truncation** unless a word is followed by a period. Placing a dollar sign (\$) at the end of a **word** allows freer **prefix matching**. For example, a search on "communicat" retrieves 642 hits, while a search on "communicat\$" retrieves...

...feature accounts for plural and singular forms of keywords ("s" and "es" are stripped). Items **retrieved** are **ranked** in order of relevance; the results list includes the document title and its relevance score...

... View the Next (10, 25, 100) Results button lets users browse results beyond the maximum **retrieval** specified. The **term** "ebola" yielded 123 hits while "pollution" found 782 hits.

WebCrawler's easy-to-use search...

15/3,K/25 (Item 25 from file: 15)
DIALOG(R)File 15:ABI/Inform(R)
(c) 2004 ProQuest Info&Learning. All rts. reserv.

00743801 93-93022
Dialog's RANK command: Building and mining the data mountain
Basch, Reva
Online v17n4 PP: 28-35 Jul 1993
ISSN: 0146-5422 JRNL CODE: ONL
WORD COUNT: 3677

...TEXT: in the top three-ranked counties.

POSSIBLE CONFUSION ALERT (PCA):

RANK defaults to the current **answer** set. To **RANK** on an earlier set, you must precede the set number with an "S," e.g...

...It will display the distinctly unhelpful message, Rank fields found in 1 records--1 unique **terms**. **Searchers** would be much better served with an error message, something like "Precede set number with..."

... specify the set number, even for the default set, but doesn't require the S- **Prefix** --RANK doesn't require the set number, but does require an "S" if you do...
? t15/3,k/27,36

15/3,K/27 (Item 27 from file: 15)
DIALOG(R)File 15:ABI/Inform(R)
(c) 2004 ProQuest Info&Learning. All rts. reserv.

00666722 93-15943
New version of QUESTEL Plus software available
Anonymous
Information Today v10n1 PP: 1, 5 Jan 1993
ISSN: 8755-6286 JRNL CODE: IFT
WORD COUNT: 556

...TEXT: it has not yet been made available. The benefit of STRINGSEARCH is the capability to **search** for strings or **terms** in a very precise order and which are not otherwise **retrievable**.

Questel's new .. **RANK** (..RK) command lets users display results in descending order of occurrence of the search terms...

15/3,K/36 (Item 6 from file: 484)
DIALOG(R)File 484:Periodical Abs Plustext
(c) 2004 ProQuest. All rts. reserv.

03830515 (USE FORMAT 7 OR 9 FOR FULLTEXT)

DejaNews and other usenet search tools

Notess, Greg R

Online (ONL), v22 n4, p74-78, p.4

Jul 1998

ISSN: 0146-5422 JOURNAL CODE: ONL

DOCUMENT TYPE: Commentary

LANGUAGE: English

RECORD TYPE: Fulltext; Abstract

WORD COUNT: 2428

TEXT:

... be m' comprehensive to use Boolean operators and NEAR in DejaNews instead of phra' searches.' ' **Truncation** is available by using an asterisk for unlimited characters and a' question mark for a...

...authors, subject words, and dates. It also gives' options for retrieving 25, 50, or 100 **hits** at a time- **sorting output** by' relevance **score** , group, author, subject, or date-and for choosing concise,' detailed, or threaded display.' ' The author...

...searching is available in both' using the field name followed by a colon and the **search term** . Available fie' are author (from:), newsgroups, subject, summary, and keywords.' ' The AltaVista display gives two...Phrase searching is available with either single or double quotes. Unless' double quotes are used, **search terms** are automatically **truncated** and **search** ' by **word stem** . For example, a **search** on epoxy also automatically searches' epoxy, epoxies, epoxied, epoxyed, and epoxys. **Search terms** and phrases can ' nested with parentheses.' ' By default, only the most recent two weeks are...

? t15/3,k/41

15/3,K/41 (Item 3 from file: 20)

DIALOG(R)File 20:Dialog Global Reporter

(c) 2004 The Dialog Corp. All rts. reserv.

06273050 (USE FORMAT 7 OR 9 FOR FULLTEXT)

Indian Search Engines: Click for the 'Home'

COMPUTERS TODAY, p72

July 31, 1999

JOURNAL CODE: WCOT LANGUAGE: English RECORD TYPE: FULLTEXT

WORD COUNT: 1278

(USE FORMAT 7 OR 9 FOR FULLTEXT)

... It also lets you use the search index, i.e. categories and sub-categories. Khoj **rates** the **matches** returned in the order of importance.

The Global Comparison

It wouldn't be fair if...

?

File 348:EUROPEAN PATENTS 1978-2004/Sep W02

(c) 2004 European Patent Office

File 349:PCT FULLTEXT 1979-2002/UB=20040916,UT=20040909

(c) 2004 WIPO/Univentio

| Set | Items | Description |
|-----|---------|---|
| S1 | 1410730 | QUERY? OR QUERIE? ? OR SUBQUER? OR SEARCH? OR FETCH? OR RETRIEV? OR TEXTSEARCH? |
| S2 | 234653 | MATCH??? ? |
| S3 | 699989 | WORD? ? OR TERM? ? |
| S4 | 7196 | CHARACTERSTRING? OR CHARACTER? ?(2N)STRING? ? OR WORDSTEM? OR MORPHEME? OR WORDELEMENT? OR BASETERM? OR BASEWORD? OR LEX-EME? |
| S5 | 16962 | S3(2N) (STEM? ? OR ELEMENT? ? OR BASE OR BASES) |
| S6 | 382575 | SUFFIX? OR PREFIX? OR DERIVATI? OR AFFIX? OR POSTFIX? OR TRUNCAT? OR LEFTTRUNCAT? OR RIGHTTRUNCAT? |
| S7 | 70442 | S1:S2(25N)S3:S5 |
| S8 | 559 | S7(25N)S6 |
| S9 | 14469 | IC='G06F-017/30':IC='G06F-017/32' |
| S10 | 4261 | IC='G06F-007' |
| S11 | 1116043 | RANK? OR RATE OR RATES OR RATED OR RATING? OR SORT??? ? OR SCOR??? ? OR VALUATION? OR TALLY? OR TALLIE? ? OR WEIGH? |
| S12 | 35099 | S2:S4(3N)S11 |
| S13 | 113896 | S11(3N) (RESULT? OR HITLIST? OR REFERENCE? OR RETRIEV? OR HIT OR HITS OR OUTPUT? OR OUT()PUT? ? OR RESPONSE? ? OR ANSWER? ? OR REPLIE? ? OR REPLY?) |
| S14 | 54756 | S1:S2(5N)S3:S5 |
| S15 | 269 | S14(25N)S6 |
| S16 | 21 | S15(25N)S12:S13 |
| ? | | |

? t16/5,k/1,4,6-7,10-13

16/5,K/1 (Item 1 from file: 348)
DIALOG(R)File 348:EUROPEAN PATENTS
(c) 2004 European Patent Office. All rts. reserv.

01626936

System and method for document retrieval
System und Verfahren fur Dokumentwiederfindung
Systeme et procede d'extraction de documents

PATENT ASSIGNEE:

Ricoh Company, Ltd., (209037), 3-6, Nakamagome 1-chome, Ohta-ku, Tokyo
143-8555, (JP), (Applicant designated States: all)

INVENTOR:

Ogawa, Yasushi, c/o Ricoh Company, Ltd., 3-6 Nakamagome 1-chome, Ohta-ku,
Tokyo 143-8555, (JP)

LEGAL REPRESENTATIVE:

Leeming, John Gerard (74731), J.A. Kemp & Co., 14 South Square, Gray's
Inn, London WC1R 5JJ, (GB)

PATENT (CC, No, Kind, Date): EP 1341101 A2 030903 (Basic)

APPLICATION (CC, No, Date): EP 2003251244 030228;

PRIORITY (CC, No, Date): JP 200253895 020228; JP 200276767 020319

DESIGNATED STATES: AT; BE; BG; CH; CY; CZ; DE; DK; EE; ES; FI; FR; GB; GR;
HU; IE; IT; LI; LU; MC; NL; PT; SE; SI; SK; TR

EXTENDED DESIGNATED STATES: AL; LT; LV; MK; RO

INTERNATIONAL PATENT CLASS: G06F-017/30

ABSTRACT EP 1341101 A2

A document retrieval system with program storage device and computer program product capable of providing excellent capabilities of accurate document retrieval even using plural indices, to thereby attain the accuracy as high as that for a single index previously employed. The document retrieval system 100 is provided with an index section 16 for storing and managing plural indices which are each generated for respective groups of documents divided to be currently retrieved; a retrieval condition analyzing section 11 for analyzing acquired retrieval conditions, dividing a retrieval character string contained in the retrieval conditions into index units, and representing the retrieval conditions in terms of a predetermined internal representation for each index; a TF computing section 12, a DF computing section 13, and a DF term computing section 14 for specifying the documents corresponding to the retrieval conditions; and a merging section 15 for merging retrieval results obtained for each index and generating final retrieval results.

ABSTRACT WORD COUNT: 156

NOTE:

Figure number on first page: 2

LEGAL STATUS (Type, Pub Date, Kind, Text):

Application: 030903 A2 Published application without search report

Examination: 030903 A2 Date of request for examination: 20030310

LANGUAGE (Publication,Procedural,Application): English; English; English

FULLTEXT AVAILABILITY:

| Available Text | Language | Update | Word Count |
|------------------------------------|-----------|--------|------------|
| CLAIMS A | (English) | 200336 | 2667 |
| SPEC A | (English) | 200336 | 18730 |
| Total word count - document A | | | 21397 |
| Total word count - document B | | | 0 |
| Total word count - documents A + B | | | 21397 |

...SPECIFICATION section 14, and to merge the thus obtained results, to thereby generate the final retrieval results (step S207), and sort the results according to the order of magnitude of the score to

subsequently output to the output unit 6 (step S208).

In case when a retrieval character string in the n-gram retrieval is found shorter than the index unit, the frequency information can be obtained by expanding the retrieval character string utilizing units prefix searched to agree with the retrieval character, and considering the document containing any of the noted...

16/5,K/4 (Item 4 from file: 348)
DIALOG(R)File 348:EUROPEAN PATENTS
(c) 2004 European Patent Office. All rts. reserv.

01188193

HANDWRITTEN OR SPOKEN WORDS RECOGNITION WITH NEURAL NETWORKS
HANDGESCHRIEBENE ODER GESPROCHENE WORT-ERKENNUNG MIT NEURONALEN NETZWERKEN
SYSTEME DE RECONNAISSANCE DE L'ECRITURE MANUSCRITE OU DE LA PAROLE
PATENT ASSIGNEE:

MICROSOFT CORPORATION, (749866), One Microsoft Way, Redmond, WA 98052,
(US), (Proprietor designated states: all)

INVENTOR:

GUHA, Angshuma, 18040 NE 109th Court, Redmond, WA 98052, (US)
HALUPTZOK, Patrick, M., 1907 E. Lake Sammamish Place SE, Issaquah, WA
98029, (US)
PITTMAN, James, A., 4064 240th Place SE, Issaquah, WA 98029, (US)

LEGAL REPRESENTATIVE:

Grunecker, Kinkeldey, Stockmair & Schwanhausser Anwaltssozietat (100721)
, Maximilianstrasse 58, 80538 Munchen, (DE)

PATENT (CC, No, Kind, Date): EP 1141941 A2 011010 (Basic)
EP 1141941 B1 030502
WO 2000041166 000713

APPLICATION (CC, No, Date): EP 99966381 991216; WO 99US30114 991216

PRIORITY (CC, No, Date): US 227170 990107

DESIGNATED STATES: AT; BE; CH; CY; DE; DK; ES; FI; FR; GB; GR; IE; IT; LI;
LU; MC; NL; PT; SE

INTERNATIONAL PATENT CLASS: G10L-015/16; G06K-009/00; G06K-009/22;
G06K-009/72

CITED PATENTS (EP B): EP 858047 A; WO 98/15914 A; US 5404422 A; US 5440651
A; US 5528728 A; US 5764797 A

CITED REFERENCES (EP B):

YU H -J ET AL: "A NEURAL NETWORK FOR 500 VOCABULARY WORD SPOTTING USING
ACOUSTIC SUB-WORD UNITS" IEEE INTERNATIONAL CONFERENCE ON ACOUSTICS,
SPEECH, AND SIGNAL PROCESSING (ICASSP),US,LOS ALAMITOS,CA: IEEE COMP.
SOC. PRESS,1997, pages 3277-3280, XP000788059 ISBN: 0-8186-7920-4

BODENHAUSEN U ET AL: "CONNECTIONIST ARCHITECTURAL LEARNING FOR HIGH
PERFORMANCE CHARACTERAND SPEECH RECOGNITION" PROCEEDINGS OF THE
INTERNATIONAL CONFERENCE ON ACOUSTICS, SPEECH, AND SIGNAL PROCESSING
(ICASSP),US,NEW YORK, IEEE, vol. 1, 27 - 30 April 1993, pages 625-628,
XP000399199 ISBN: 0-7803-0946-4

SHI H ET AL: "LEXICON-DRIVEN HANDWRITTEN WORD RECOGNITION USING CHOQUET
FUZZY INTEGRAL" IEEE INTERNATIONAL CONFERENCE ON SYSTEMS, MAN AND
CYBERNETICS,US,NEW YORK, IEEE, 1996, pages 412-417, XP000729948 ISBN:
0-7803-3281-4

BENGIO Y ET AL: "LeRec: a NN/HMM hybrid for on-line handwriting
recognition" NEURAL COMPUTATION, NOV. 1995, USA, vol. 7, no. 6, pages
1289-1303, XP000789843 ISSN: 0899-7667;

NOTE:

No A-document published by EPO

LEGAL STATUS (Type, Pub Date, Kind, Text):

Application: 000906 A2 International application. (Art. 158(1))

Application: 000906 A2 International application entering European

phase

Application: 011010 A2 Published application without search report

Examination: 011010 A2 Date of request for examination: 20010704

Examination: 011205 A2 Date of dispatch of the first examination report: 20011022

Change: 020911 A2 International Patent Classification changed: 20020722

Change: 020911 A2 Title of invention (French) changed: 20020722

Grant: 030502 B1 Granted patent

Lapse: 031210 B1 Date of lapse of European Patent in a contracting state (Country, date): SE 20030802,

Lapse: 040114 B1 Date of lapse of European Patent in a contracting state (Country, date): AT 20030502, CH 20030502, LI 20030502, ES 20030813, GR 20030802, NL 20030502, PT 20030804, SE 20030802,

Lapse: 040128 B1 Date of lapse of European Patent in a contracting state (Country, date): AT 20030502, CH 20030502, LI 20030502, ES 20030813, FI 20030502, GR 20030802, NL 20030502, PT 20030804, SE 20030802,

Oppn None: 040421 B1 No opposition filed: 20040203

Lapse: 040602 B1 Date of lapse of European Patent in a contracting state (Country, date): AT 20030502, CH 20030502, LI 20030502, DK 20030802, ES 20030813, FI 20030502, GR 20030802, NL 20030502, PT 20030804, SE 20030802,

Lapse: 040707 B1 Date of lapse of European Patent in a contracting state (Country, date): AT 20030502, BE 20030502, CH 20030502, LI 20030502, DK 20030802, ES 20030813, FI 20030502, GR 20030802, NL 20030502, PT 20030804, SE 20030802,

Lapse: 040901 B1 Date of lapse of European Patent in a contracting state (Country, date): AT 20030502, BE 20030502, CH 20030502, LI 20030502, CY 20031216, DK 20030802, ES 20030813, FI 20030502, GR 20030802, NL 20030502, PT 20030804, SE 20030802,

LANGUAGE (Publication,Procedural,Application): English; English; English

FULLTEXT AVAILABILITY:

| Available Text | Language | Update | Word Count |
|------------------------------------|-----------|--------|------------|
| CLAIMS B | (English) | 200318 | 927 |
| CLAIMS B | (German) | 200318 | 970 |
| CLAIMS B | (French) | 200318 | 1091 |
| SPEC B | (English) | 200318 | 6673 |
| Total word count - document A | | | 0 |
| Total word count - document B | | | 9661 |
| Total word count - documents A + B | | | 9661 |

...SPECIFICATION DTW) 82 to find the most probable path through the output matrix 80 for that word . Then, the score for each word may be compared with the scores of other words to find the best word or words . Note that searching for the least expensive word in this manner is greatly sped up by using a trie structured dictionary 84 (FIG. 2) so that the cost of each common prefix is only computed once, whereby only the ending letters that are different need their costs recomputed. Also, thresholding the search against the score for the best word found up to that point prevents searches down futile paths,

while searching down the most...

16/5,K/6 (Item 6 from file: 348)
DIALOG(R)File 348:EUROPEAN PATENTS
(c) 2004 European Patent Office. All rts. reserv.

00699063

Information retrieval method.

Informationswiederauffindungsverfahren.

Procede de recouvrement d'informations.

PATENT ASSIGNEE:

AT&T Corp., (589370), 32 Avenue of the Americas, New York, NY 10013-2412,
(US), (applicant designated states: DE;FR;GB)

INVENTOR:

Church, Kenneth Ward, 2 Longhill Lane, Chatham, New Jersey 07928, (US)
Helfman, Jonathan Isaac, 151 Riverview Avenue, Gillette, New Jersey 07933
, (US)
Lewis, David Dolan, 851 Springfield Avenue, Apt. 10G, Summit, New Jersey
07901, (US)

LEGAL REPRESENTATIVE:

Watts, Christopher Malcolm Kelway, Dr. et al (37391), AT&T (UK) Ltd. 5,
Morningson Road, Woodford Green Essex, IG8 0TU, (GB)

PATENT (CC, No, Kind, Date): EP 665504 A1 950802 (Basic)

APPLICATION (CC, No, Date): EP 95300288 950118;

PRIORITY (CC, No, Date): US 188002 940128

DESIGNATED STATES: DE; FR; GB

INTERNATIONAL PATENT CLASS: G06F-017/30;

ABSTRACT EP 665504 A1

An information retrieval method wherein users may submit a query via a graphical bitmapping technique. The user provides an information retrieval system with a bitmap of a printed, written, or graphical query by either scanning the query with a graphical scanner, or employing a standard facsimile transmission machine. The information retrieval system then performs an optical image/character recognition process upon the received bitmap to determine the content of the query, information is then retried based upon the recognized characters and images. In a particular method of the invention, the user is provided with a bitmap of the retrieved information. (see image in original document)

ABSTRACT WORD COUNT: 106

LEGAL STATUS (Type, Pub Date, Kind, Text):

Application: 950802 A1 Published application (A1with Search Report
;A2without Search Report)

Examination: 960320 A1 Date of filing of request for examination:
960117

Withdrawal: 971229 A1 Date on which the European patent application
was withdrawn: 971103

LANGUAGE (Publication,Procedural,Application): English; English; English

FULLTEXT AVAILABILITY:

| Available Text | Language | Update | Word Count |
|------------------------------------|-----------|--------|------------|
| CLAIMS A | (English) | EPAB95 | 229 |
| SPEC A | (English) | EPAB95 | 1870 |
| Total word count - document A | | | 2099 |
| Total word count - document B | | | 0 |
| Total word count - documents A + B | | | 2099 |

...SPECIFICATION of the bitmap in this example would include the recognition of text characters and the derivation of particular search parameters by application of term weighting techniques. The

information retrieved by the search is then transmitted back to the requesting user.

In any of the...

16/5,K/7 (Item 7 from file: 348)

DIALOG(R) File 348:EUROPEAN PATENTS

(c) 2004 European Patent Office. All rts. reserv.

00285369

Paradigm-based morphological text analysis for natural languages.

Auf Paradigmen basierende morphologische Textanalyse für natürliche Sprachen.

Analyse morphologique de textes pour des langues naturelles basée sur paradigmes.

PATENT ASSIGNEE:

International Business Machines Corporation, (200120), Old Orchard Road, Armonk, N.Y. 10504, (US), (applicant designated states: DE;FR;GB)

INVENTOR:

Zamora, Antonio, 4601 North Park Avenue, Chevy Chase Maryland 20815, (US)

LEGAL REPRESENTATIVE:

Teufel, Fritz, Dipl.-Phys. (11855), IBM Deutschland Informationssysteme GmbH, Patentwesen und Urheberrecht, D-70548 Stuttgart, (DE)

PATENT (CC, No, Kind, Date): EP 282721 A2 880921 (Basic)

EP 282721 A3 900627

EP 282721 B1 950531

APPLICATION (CC, No, Date): EP 88101694 880205;

PRIORITY (CC, No, Date): US 28437 870320

DESIGNATED STATES: DE; FR; GB

INTERNATIONAL PATENT CLASS: G06F-017/20; G06F-017/27;

CITED PATENTS (EP A): FR 2569883 A; EP 12777 A; US 4495566 A; EP 266001 A; EP 241717 A

CITED REFERENCES (EP A):

IBM TECHNICAL DISCLOSURE BULLETIN, vol. 26, no. 11, April 1984, pages 6085-6086, Armonk, New York, US; R.G. CARLGREN: "Language-independent dictionary storage and access technique";

ABSTRACT EP 282721 A2

A computer method is disclosed for analyzing text by employing a model known as a paradigm, that provides all the inflectional forms of a word. A file structure is created consisting of two components, a list of words (a dictionary), each word of which is associated with a set of paradigm references, and the file of paradigms consisting of grammatical categories paired with their corresponding ending or affix portions (known as the desinence) specifying tense, mood, number, gender or other linguistic attribute. A computer method is disclosed for generating the file structure of the dictionary by generating all forms of the words from a list of standard forms of the words (known as the lemma) which is generally the infinitive of a verb of the singular form of a noun, the lemmas being generated with their corresponding paradigms. The method sorts and organizes the resulting word list into a dictionary. An input data stream of natural language words can then be processed by generating a lemma for each input word. The specific grammatical form of an input word can be generated from the standard form of the word (the lemma) and the grammatical category, by matching the lemma against the dictionary and using its paradigm references to access a set of paradigms. Then the desinences of the paradigms are matched against the lemma and the desinence corresponding to the specified grammatical category is selected. The specific grammatical form is generated by replacing the desinence of the lemma with the desinence of the desired grammatical form.

ABSTRACT WORD COUNT: 262

LEGAL STATUS (Type, Pub Date, Kind, Text):

Application: 880921 A2 Published application (Alwith Search Report
;A2without Search Report)
Examination: 890315 A2 Date of filing of request for examination:
890117
Change: 900620 A2 Obligatory supplementary classification
(change)
Search Report: 900627 A3 Separate publication of the European or
International search report
Examination: 921209 A2 Date of despatch of first examination report:
921022
Grant: 950531 B1 Granted patent
Oppn None: 960522 B1 No opposition filed
Lapse: 970402 B1 Date of lapse of the European patent in a
Contracting State: GB 960205
Lapse: 970423 B1 Date of lapse of the European patent in a
Contracting State: DE 961101, GB 960205

LANGUAGE (Publication,Procedural,Application): English; English; English
FULLTEXT AVAILABILITY:

| Available Text | Language | Update | Word Count |
|------------------------------------|-----------|--------|------------|
| CLAIMS A | (English) | EPABF1 | 324 |
| CLAIMS B | (English) | EPAB95 | 242 |
| CLAIMS B | (German) | EPAB95 | 243 |
| CLAIMS B | (French) | EPAB95 | 286 |
| SPEC A | (English) | EPABF1 | 3385 |
| SPEC B | (English) | EPAB95 | 3558 |
| Total word count - document A | | | 3709 |
| Total word count - document B | | | 4329 |
| Total word count - documents A + B | | | 8038 |

...SPECIFICATION singular form of a noun, the lemmas being generated with their corresponding paradigms. The method **sorts** and organizes the **resulting** word list into a dictionary. An input data stream of natural language words can then be processed by generating a lemma for each input **word**. This is done by **matching** the input **word** against the dictionary and using the resulting paradigm references to access a set of paradigms. Then the ending or **affix** (desinence) of the paradigm is **matched** against the input **word** and the corresponding grammatical category for each matched desinence is recorded and the standard form...

...SPECIFICATION singular form of a noun, the lemmas being generated with their corresponding paradigms. The method **sorts** and organizes the **resulting** word list into a dictionary. An input data stream of natural language words can then be processed by generating a lemma for each input **word**. This is done by **matching** the input **word** against the dictionary and using the resulting paradigm references to access a set of paradigms. Then the ending or **affix** (desinence) of the paradigm is **matched** against the input **word** and the corresponding grammatical category for each matched desinence is recorded and the standard form...

16/5,K/10 (Item 3 from file: 349)
DIALOG(R)File 349:PCT FULLTEXT
(c) 2004 WIPO/Univentio. All rts. reserv.

00888749 **Image available**

PLANT GENES, THE EXPRESSION OF WHICH ARE ALTERED BY PATHOGEN INFECTION
GENES DE PLANTES DONT L'EXPRESSION EST MODIFIEE PAR L'INFECTION PAR UN

PATHOGENE

Patent Applicant/Assignee:

SYNGENTA PARTICIPATIONS AG, Schwarzwaldallee 215, CH-4058 Basel, CH, CH
(Residence), CH (Nationality), (For all designated states except: US)
UNIVERSITY OF NORTH CAROLINA AT CHAPEL HILL, 300 Bynum Hall, Campus Box
4100, Chapel Hill, NC 27599-4100, US, US (Residence), US (Nationality),
(For all designated states except: US)

Patent Applicant/Inventor:

GLAZEBROOK Jane, 4503 Ocean Valley Lane, San Diego, CA 92130, US, US
(Residence), US (Nationality)
WANG Xun, 12524 Caminito Vista Soledad, San Diego, CA 92121, US, US
(Residence), -- (Nationality)
DANGL Jeffrey L, 601 Jones Ferry Road, Apt. B, Carrboro, NC 27510, US, US
(Residence), US (Nationality)
EULGEM Thomas, 605 Jones Ferry Road, Apt. VV1, Carrboro, NC 27510, US, US
(Residence), US (Nationality)
ZHU Tong, 5260 Caminito Exquisito, San Diego, CA 92130, US, US
(Residence), -- (Nationality)

Legal Representative:

VIKSNINS Ann S (agent), Schwegman, Lunberg, Woessner & Kluth, P.O. Box
2938, Minneapolis, MN 55402, US,

Patent and Priority Information (Country, Number, Date):

Patent: WO 200222675 A2-A3 20020321 (WO 0222675)
Application: WO 2001US28506 20010914 (PCT/WO US0128506)
Priority Application: US 2000232778 20000915; US 2001300183 20010622

Designated States:

(Protection type is "patent" unless otherwise stated - for applications
prior to 2004)

AE AG AL AM AT AU AZ BA BB BG BR BY BZ CA CH CN CO CR CU CZ DE DK DM DZ
EC EE ES FI GB GD GE GH GM HR HU ID IL IN IS JP KE KG KP KR KZ LC LK LR
LS LT LU LV MA MD MG MK MN MW MX MZ NO NZ PH PL PT RO RU SD SE SG SI SK
SL TJ TM TR TT TZ UA UG US UZ VN YU ZA ZW
(EP) AT BE CH CY DE DK ES FI FR GB GR IE IT LU MC NL PT SE TR
(OA) BF BJ CF CG CI CM GA GN GQ GW ML MR NE SN TD TG
(AP) GH GM KE LS MW MZ SD SL SZ TZ UG ZW
(EA) AM AZ BY KG KZ MD RU TJ TM

Main International Patent Class: C12N-015/82

International Patent Class: C12Q-001/68; A01H-005/00; A01H-005/10;
C07K-014/37; C07K-014/415

Publication Language: English

Filing Language: English

Fulltext Availability:

Detailed Description

Claims

Fulltext Word Count: 284445

English Abstract

Methods to identify genes, the expression of which is altered in response
to pathogen infection, are provided, as well as the genes identified
thereby.

French Abstract

L'invention concerne des procedes d'identification de genes dont
l'expression est modifiee en reponse a l'infection par un pathogene,
ainsi que les genes identifies de cette maniere.

Legal Status (Type, Date, Text)

Publication 20020321 A2 Without international search report and to be
republished upon receipt of that report.

Publication 20020321 A2 Sequence listing published separately in
electronic form and available upon request from the

International Bureau.
Examination 20021114 Request for preliminary examination prior to end of
19th month from priority date
Search Rpt 20030710 Late publication of international search report
Republication 20030710 A3 With international search report.
Republication 20030710 A3 Sequence listing published separately in
electronic form and available upon request from the
International Bureau.

Fulltext Availability:
Detailed Description

Detailed Description

... acid (RNA) is involved in
the transfer of information contained within DNA into proteins. The term

"nucleotide sequence" refers to a polymer of DNA or RNA which can be
single or double...

...non-natural or altered
nucleotide bases capable of incorporation into DNA or RNA polymers. The
terms "nucleic acid", "nucleic acid molecule", "nucleic acid fragment"
or "nucleic acid sequence or segment" may...nih.gov/). This algorithm
involves first identifying high scoring sequence pairs (HSPs) by
identifying short words of length W in the query sequence, which
either match or satisfy some positive-valued threshold score T when
aligned with...

...of the same length in a database sequence. T is referred to as
the-neighborhood word score threshold (Altschul et al., 1990).

These initial neighborhood word hits act as seeds for initiating
searches to find

34

longer HSPs containing them. The word hits are then extended in both
directions

along each sequence for as far as the cumulative...

16/5,K/11 (Item 4 from file: 349)
DIALOG(R)File 349:PCT FULLTEXT
(c) 2004 WIPO/Univentio. All rts. reserv.

00852426 **Image available**

IDENTIFYING CLUSTERS OF TRANSCRIPTION FACTOR BINDING SITES

IDENTIFICATION DE GROUPES DE SITES DE LIAISON DU FACTEUR DE TRANSCRIPTION

Patent Applicant/Assignee:

WASHINGTON UNIVERSITY, One Brookings Drive, St. Louis, MO 63130, US, US
(Residence), US (Nationality)

Inventor(s):

STATES David J, 4543 Pershing Place, St. Louis, MO 63108, US,

Legal Representative:

BAIN Robert M (et al) (agent), Senniger, Powers, Leavitt & Roedel, One
Metropolitan Square, 16th Floor, St. Louis, MO 63102, US,

Patent and Priority Information (Country, Number, Date):

Patent: WO 200185915 A2-A3 20011115 (WO 0185915)

Application: WO 2001US15291 20010511 (PCT/WO US0115291)

Priority Application: US 2000203469 20000511; US 2001853141 20010510

Designated States:

(Protection type is "patent" unless otherwise stated - for applications

prior to 2004)

AE AG AL AM AT AU AZ BA BB BG BR BY BZ CA CH CN CR CU CZ DE DK DM DZ EE
ES FI GB GD GE GH GM HR HU ID IL IN IS JP KE KG KP KR KZ LC LK LR LS LT
LU LV MA MD MG MK MN MW MX MZ NO NZ PL PT RO RU SD SE SG SI SK SL TJ TM
TR TT TZ UA UG UZ VN YU ZA ZW

(EP) AT BE CH CY DE DK ES FI FR GB GR IE IT LU MC NL PT SE TR

(OA) BF BJ CF CG CI CM GA GN GW ML MR NE SN TD TG

(AP) GH GM KE LS MW MZ SD SL SZ TZ UG ZW

(EA) AM AZ BY KG KZ MD RU TJ TM

Main International Patent Class: G06F-019/00

International Patent Class: G01N-033/483

Publication Language: English

Filing Language: English

Fulltext Availability:

Detailed Description

Claims

Fulltext Word Count: 8431

English Abstract

Identifying clusters of protein binding sites in a nucleotide sequence under analysis. A computerized system determines likelihood parameters for a plurality of known protein binding sites. The likelihood parameter for each protein binding site represents a likelihood that the protein binding site will occur in a nucleotide sequence under analysis relative to a likelihood that the protein binding site will occur in a random nucleotide sequence of a substantially equivalent composition. Selected protein binding sites are grouped as a function of their respective likelihood parameters to determine a likelihood score, which is compared to a predetermined threshold. The selected protein binding sites in the nucleotide sequence are identified as one or more clusters if the likelihood score exceeds the predetermined threshold.

French Abstract

La presente invention concerne l'identification de groupes de liaison de proteines dans une sequence nucleotidique analysee. Un systeme informatise determine des parametres de vraisemblance pour une pluralite de sites de liaison de proteines connus. Le parametre de vraisemblance pour chaque site de liaison de proteine represente la vraisemblance que le site de liaison de proteine se trouve dans une sequence nucleotidique analysee par rapport a la vraisemblance que le site de liaison de proteine se situe dans une sequence nucleotidique aleatoire d'une composition sensiblement equivalente. Des sites de liaison de proteines selectionnes sont groupes en fonction de leurs parametres de vraisemblance respectifs pour determiner une valeur de vraisemblance qui est comparee a un seuil predefini. Les sites de liaison de proteines selectionnes presents dans la sequence nucleotidique sont identifies comme formant un ou plusieurs groupes si la valeur de vraisemblance depasse le seuil predefini.

Legal Status (Type, Date, Text)

Publication 20011115 A2 Without international search report and to be republished upon receipt of that report.

Search Rpt 20020307 Late publication of international search report

Republication 20020307 A3 With international search report.

Examination 20020627 Request for preliminary examination prior to end of 19th month from priority date

Fulltext Availability:

Detailed Description

Detailed Description

... is less than 10

nucleotides, the pattern is extended to 10 nucleotides by

including all **suffix** strings with zero score. To search a query sequence, incremental segments of 10 characters are used to generate a **search word**, and this **word** is used to look up potential hits in the index. The full pattern **score** for each candidate **hit** is then evaluated explicitly. Note that this algorithm finds all pattern hits scoring above C...

16/5,K/12 (Item 5 from file: 349)
DIALOG(R)File 349:PCT FULLTEXT
(c) 2004 WIPO/Univentio. All rts. reserv.

00846371 **Image available**

**METHOD AND SYSTEM FOR RETRIEVING INFORMATION BASED ON MEANINGFUL CORE WORD
PROCEDE ET SYSTEME POUR L'EXTRACTION D'INFORMATIONS EN FONCTION D'UN MOT
CENTRAL SIGNIFIANT**

Patent Applicant/Assignee:

KOREA TELECOM, 206, Jungja-dong, Pundang-ku, Sungnam-shi, Kyoungki-do
463-010, KR, KR (Residence), KR (Nationality), (For all designated
states except: US)

Patent Applicant/Inventor:

JUNG Il-Hyung, 112-1503 SeonsaHyundai Apt., Amsa 2-dong, Kangdong-gu,
Seoul 134-052, KR, KR (Residence), KR (Nationality), (Designated only
for: US)

Legal Representative:

SHINSUNG PATENT FIRM (agent), Haechon Building, 741-40, Yeoksaml-dong,
Kangnam-ku, Seoul 135-081, KR,

Patent and Priority Information (Country, Number, Date):

Patent: WO 200180077 A1 20011025 (WO 0180077)

Application: WO 2001KR650 20010418 (PCT/WO KR0100650)

Priority Application: KR 200020398 20000418

Designated States:

(Protection type is "patent" unless otherwise stated - for applications
prior to 2004)

AE AG AL AM AT AU AZ BA BB BG BR BY BZ CA CH CN CR CU CZ DE DK DM DZ EE
ES FI GB GD GE GH GM HR HU ID IL IN IS JP KE KG KP KZ LC LK LR LS LT LU
LV MA MD MG MK MN MW MX MZ NO NZ PL PT RO RU SD SE SG SI SK SL TJ TM TR
TT TZ UA UG US UZ VN YU ZA ZW

(EP) AT BE CH CY DE DK ES FI FR GB GR IE IT LU MC NL PT SE TR

(OA) BF BJ CF CG CI CM GA GN GW ML MR NE SN TD TG

(AP) GH GM KE LS MW MZ SD SL SZ TZ UG ZW

(EA) AM AZ BY KG KZ MD RU TJ TM

Main International Patent Class: G06F-017/30

Publication Language: English

Filing Language: Korean

Fulltext Availability:

Detailed Description

Claims

Fulltext Word Count: 10130

English Abstract

The present invention relates to a method and system for extracting a meaningful core word from a query and a method and system for retrieving information based on the same are disclosed. The system for retrieving extracts a meaningful core word of a lemma, expands the lemma and retrieves texts based on the expanded lemma, to thereby improve performance of the retrieval system and convenience of a user.

French Abstract

L'invention concerne un procede et un systeme pour l'extraction d'un mot central signifiant d'une demande, ainsi qu'un procede et un systeme pour

extraire des informations en fonction dudit mot central. Ledit systeme d'extraction extrait un mot central signifiant d'un lemme, procede a l'extension du lemme et extrait des textes en fonction du lemme etendu, ce qui ameliore ainsi les performances du systeme d'extraction et la convivialite pour l'utilisateur.

Legal Status (Type, Date, Text)

Publication 20011025 A1 With international search report.

Examination 20020207 Request for preliminary examination prior to end of 19th month from priority date

Fulltext Availability:

Detailed Description

Detailed Description

... a lemma for accessing to the core word dictionary 23, extracting words, stem words or **derivatives**, having core meaning of the lemma and conducting search with the lemma set above or extracted stem words or **derivative** as a key word for **searching** after expanding the lemma, and an result output unit 24 which puts different **weights** on the key words before expansion(lemmas) and key words after expansion(stem words or derivatives) - that is, putting different **weights** on the **results** acquired by using a lemma as a key word and ones by using a stem word or **derivative** as a key word - and outputs **search** results in the priority order by the weight.

In case that the core word dictionary...

...the corresponding lemma. In this case, the core word dictionary 23 can be constructed putting **weights** on the stem word or **derivative** in advance while being constructed, Thus, all you need to do is output the results **searched** with corresponding stem word or **derivative** in a corresponding order. Meanwhile, the information retrieval system described above needs the steps of...extracted derivative as a key word, After that, the result output unit 24 puts different **weights** on the key word before expansion (lemma) and the key word after expansion (stem word or **derivative**).

In other words, different **weights** are put on the result searched with the lemma as a key word and on the one **searched** with the stem word or **derivative** as a key word, Then at step 407,, the **search** results are outputted to the user in the priority order according to weight. In the...

16/5,K/13 (Item 6 from file: 349)

DIALOG(R)File 349:PCT FULLTEXT

(c) 2004 WIPO/Univentio. All rts. reserv.

00845260

METHOD AND SYSTEM FOR CONDUCTING A FULL TEXT SEARCH ON A CLIENT SYSTEM BY A SERVER SYSTEM

PROCEDE ET SYSTEME POUR EFFECTUER UNE RECHERCHE PLEIN TEXTE SUR UN SYSTEME CLIENT PAR UN SYSTEME SERVEUR

Patent Applicant/Assignee:

GLOBALSCAPE INC, Suite 101, 6000 NW Parkway, San Antonio, TX 78249, US,
US (Residence), US (Nationality)

Inventor(s):

CHRISTAL David B, 18910 Red River Pass, San Antonio, TX 78259, US,
KHATRI Nimesh, Apt. 303, 7750 Pipes Lane, San Antonio, TX 78251, US,

Legal Representative:

KAMMER Mark A (et al) (agent), Cox & Smith Incorporated, 112 East Pecan
Street, Suite 1800, San Antonio, TX 78205, US,

Patent and Priority Information (Country, Number, Date):

Patent: WO 200177898 A1 20011018 (WO 0177898)

Application: WO 2001US10702 20010404 (PCT/WO US0110702)

Priority Application: US 2000194428 20000404

Designated States:

(Protection type is "patent" unless otherwise stated - for applications
prior to 2004)

AE AG AL AM AT AU AZ BA BB BG BR BY BZ CA CH CN CO CR CU CZ DE DK DM DZ
EE ES FI GB GD GE GH GM HR HU ID IL IN IS JP KE KG KP KR KZ LC LK LR LS
LT LU LV MA MD MG MK MN MW MX MZ NO NZ PL PT RO RU SD SE SG SI SK SL TJ
TM TR TT TZ UA UG UZ VN YU ZA ZW

(EP) AT BE CH CY DE DK ES FI FR GB GR IE IT LU MC NL PT SE TR

(OA) BF BJ CF CG CI CM GA GN GW ML MR NE SN TD TG

(AP) GH GM KE LS MW MZ SD SL SZ TZ UG ZW

(EA) AM AZ BY KG KZ MD RU TJ TM

Main International Patent Class: G06F-017/30

International Patent Class: G06F-007/00

Publication Language: English

Filing Language: English

Fulltext Availability:

Detailed Description

Claims

Fulltext Word Count: 7662

English Abstract

A system and method for conducting a full text search on a client system by creating a full text search index of a string of characters on the client system for use on a server system. When the client system signs on to a server system, the client's system searches for relevant data and file information that the user is willing to share and creates a string of characters that contains information such as file name, location and size. A second client system signing on to the server system can initiate a search of the memory of the server for a selected sub-string of characters. Once the selected sub-string of characters is found, the server system sends the second client system a list of the located relevant information. If the second user wants to obtain a copy of the data, a message is sent directly between the second client and the first client system without the server system being involved unless the first client is behind a firewall. If the first client is behind a firewall, the request for the file is relayed through the server system. The requested data will then be transferred from the first client system to the second client system. Each time a client signs on, a new string of characters and suffix array is generated thus enabling the server system to be able to provide a dynamic and constantly updated index of data available for transfer between client systems.

French Abstract

L'invention concerne un procede et un systeme pour effectuer une recherche plein texte sur un systeme client par creation d'un index de recherche plein texte d'une chaine de caracteres sur le systeme client destine a etre utilise sur un systeme serveur. Lorsque le systeme client demande a se connecter sur le systeme serveur, le systeme client cherche les donnees et les informations de fichier pertinentes que l'utilisateur

souhaite partager, et cree une chaine de caracteres qui contient des informations telles que le nom, l'emplacement et la taille du fichier. Un second systeme client demandant a se connecter au systeme serveur peut lancer une recherche d'une sous-chaine de caracteres selectionnee dans la memoire du serveur. Une fois cette sous-chaine trouvee, le systeme serveur envoie au second systeme client une liste d'informations pertinentes definies. Si le second utilisateur souhaite obtenir une copie des donnees, un message est envoye directement entre le second et le premier systeme client, sans que le systeme serveur ne soit implique, a moins que le premier client se trouve derriere un pare-feu. Dans ce cas, la demande de fichier est relayee par le systeme serveur. Les donnees requises sont alors transferees du premier au second systeme client. A chaque fois qu'un client demande a se connecter, une nouvelle chaine de caracteres et un ensemble de suffixes sont produits, ce qui permet au systeme serveur de fournir un index, dynamique et constamment mis a jour, de donnees disponibles pour le transfert entre les systemes clients.

Legal Status (Type, Date, Text)

Publication 20011018 A1 With international search report.

Publication 20011018 A1 Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.

Claim Mod 20020214 Later publication of amended claims under Article 19 received: 20011022

Republication 20020214 A1 With international search report.

Republication 20020214 A1 With amended claims.

Examination 20020404 Request for preliminary examination prior to end of 19th month from priority date

Fulltext Availability:

Detailed Description

Detailed Description

... suffix array (1, 3, 5, 01, 2,r 4,r 6,r 7) . The sorted **suffix** array is the index in the preferred embodiment for rapidly and efficiently searching the original string "bananas".

The **search** server stores the original **string** of **characters** and **sorted suffix** array in memory.

Example 3

The following is an example of how a binary search...

? t16/5,k/16-17

16/5,K/16 (Item 9 from file: 349)

DIALOG(R)File 349:PCT FULLTEXT

(c) 2004 WIPO/Univentio. All rts. reserv.

00539942 **Image available**

A SEARCH SYSTEM AND METHOD FOR RETRIEVAL OF DATA, AND THE USE THEREOF IN A SEARCH ENGINE

SYSTEME ET PROCEDE EN VUE DE LA RECUPERATION DE DONNEES ET SON UTILISATION DANS UN AUTOMATE DE RECHERCHE

Patent Applicant/Assignee:

FAST SEARCH & TRANSFER ASA,

RISVIK Knut Magne,

Inventor(s):

RISVIK Knut Magne,

Patent and Priority Information (Country, Number, Date):

Patent: WO 200003315 A2 20000120 (WO 0003315)
Application: WO 99NO233 19990709 (PCT/WO NO9900233)
Priority Application: NO 983175 19980710

Designated States:

(Protection type is "patent" unless otherwise stated - for applications prior to 2004)

AE AL AM AT AU AZ BA BB BG BR BY CA CH CN CU CZ DE DK EE ES FI GB GD GE
GH HR HU ID IL IN IS JP KE KG KP KR KZ LC LK LR LS LT LU LV MD MG MK MN
MW MX NO NZ PL PT RO RU SD SE SG SI SK SL TJ TM TR TT UA UG US UZ VN YU
ZA ZW GH GM KE LS MW SD SL SZ UG ZW AM AZ BY KG KZ MD RU TJ TM AT BE CH
CY DE DK ES FI FR GB GR IE IT LU MC NL PT SE BF BJ CF CG CI CM GA GN GW
ML MR NE SN TD TG

Main International Patent Class: G06F-017/30

Publication Language: English

English Abstract

A search system for information retrieval comprises a data structure for storing a text T, a combined metric M which includes an edit distance metric for approximate degree of matching between words and/or symbols or sequences thereof in the text T and words and/or symbols in a sequence P, weighting cost functions for edit operations which transform a sequence S of words or symbols into the sequence P, and a search algorithm for determining the degree of matching between words or word sequences in a suffix tree representation of respectively the text T and a query Q. The algorithm searches the data structure with the query Q, retrieving information with specified match to the query. A method in a search system for information **retrieval** generates a **word**-spaced sparse **suffix** tree for storing **suffixes** of words in a text T as word sequence information, and a word size-dependent edit distance metric for word sequences S, P and including **word-weighted** cost functions for edit operations, and determines matches between word sequences SR or retrieved information R and word sequences PQ of a query Q by calculating the edit distance for all matches. Use in an approximate search engine.

French Abstract

L'invention porte sur un systeme de recherche d'informations comportant; une structure de donnees de stockage d'un texte T; un metrique M combine de mesure du niveau de concordance evaluant le niveau approximatif de concordance entre des mots et/ou des symboles, ou des phrases en etant faites, du texte T, et des mots ou symboles d'une sequence P; des fonctions de ponderation des mots dans des operations de mise au point de textes transformant une sequence S de mots ou de symboles en une sequence P; et un algorithme de recherche determinant le niveau de concordance entre des mots ou sequences de mots dans une representation presentant respectivement le texte T et la question Q. L'algorithme recherche la structure de donnees en posant la question Q et recupere l'information correspondant specifiquement. L'invention porte en outre sur un procede lie a un systeme de recherche d'informations produisant un arbre a suffixe de mots clairsemees stockant des suffixes de mots d'un texte T sous forme d'une sequence de mots d'information, recourant a un metrique de mesure du niveau de concordance entre les sequences de mots S et P, comportant des fonctions de cout ponderees en mots pour les operations de mise au point de textes, et determinant les correspondances entre les sequences de mots Sr des informations R recuperees et les sequences de mots Pq de la demande Q en calculant le niveau de concordance pour toutes les correspondances. L'invention porte en outre sur son utilisation dans un automate de recherche par approximation.

English Abstract

...information with specified match to the query. A method in a search system for information **retrieval** generates a **word**-spaced sparse

suffix tree for storing suffixes of words in a text T as word sequence information, and a word size-dependent edit distance metric for word sequences S, P and including word - weighted cost functions for edit operations, and determines matches between word sequences SR or retrieved information...

16/5,K/17 (Item 10 from file: 349)
DIALOG(R)File 349:PCT FULLTEXT
(c) 2004 WIPO/Univentio. All rts. reserv.

00467879 **Image available**

TEXT CLASSIFICATION SYSTEM AND METHOD
SYSTEME ET PROCEDE DE CLASSIFICATION DE TEXTES

Patent Applicant/Assignee:

THE DIALOG CORPORATION,

Inventor(s):

ZHILYAEV Maxim,

Patent and Priority Information (Country, Number, Date):

Patent: WO 9858344 A1 19981223

Application: WO 98US12604 19980616 (PCT/WO US9812604)

Priority Application: US 97876271 19970616

Designated States:

(Protection type is "patent" unless otherwise stated - for applications prior to 2004)

AL AM AT AU AZ BA BB BG BR BY CA CH CN CU CZ DE DK EE ES FI GB GE GH HU
IL IS JP KE KG KP KR KZ LC LK LR LS LT LU LV MD MG MK MN MW MX NO NZ PL
PT RO RU SD SE SG SI SK SL TJ TM TR TT UA UG UZ VN YU ZW GH GM KE LS MW
SD SZ UG ZW AM AZ BY KG KZ MD RU TJ TM AT BE CH CY DE DK ES FI FR GB GR
IE IT LU MC NL PT SE BF BJ CF CG CI CM GA GN ML MR NE SN TD TG

Main International Patent Class: G06K-009/62

International Patent Class: G06K-09:68; G06K-09:70; G06K-09:74

Publication Language: English

Fulltext Availability:

Detailed Description

Claims

Fulltext Word Count: 12259

English Abstract

Documents are classified (20) into one or more clusters (C) corresponding to predefined classification categories by building a knowledge base (22) comprising matrices of vectors which indicate the significance of terms within a corpus (T) of text formed by the documents and classified (20) in the knowledge base (22) to each cluster (C). The significance of terms is determined assuming a standard normal probability distribution, and terms are determined to be significant to a cluster if their probability of occurrence being due to chance is low. For each cluster, statistical signatures comprising sums of weighted products and intersections of cluster terms to corpus (T) terms are generated and used as discriminators for classifying documents. The knowledge base (22) is built using prefix and suffix lexical rules (38) which are context-sensitive and applied selectively to improve the accuracy and precision of classification.

French Abstract

Des documents sont classifiés (20) sous forme d'une ou plusieurs grappes (C) correspondant à des catégories de classification prédéfinies, par la construction d'une base de connaissances (22) comprenant des matrices de vecteurs qui indiquent la signification de termes au sein d'un corpus (T) de textes formé par les documents et classifié (20), dans la base de connaissances (22), en fonction de chaque grappe (C). La signification de termes est déterminée par l'adoption d'une répartition statistique

normale standard, et les termes sont designes comme etant significatifs par rapport a une grappe si leur probabilite d'apparition due au hasard est faible. Pour chaque grappe, des signatures statistiques comprenant les sommes de produits ponderes et des intersection de termes de grappe avec les termes du corpus (T) sont generees et utilisees comme discriminateurs pour les documents classifies. La base de connaissances (22) est construite au moyen de regles lexicales prefixes et suffixes (38) qui sont sensibles au contexte et sont appliquees selectivement pour l'amelioration de la precision et de l'exactitude de la classification.

Fulltext Availability:

Claims

Claim

... which have a similarity measure greater than a predetermined value; for each word, generating candidate **suffix** and **prefix** lexical rules by **matching character strings** in portions of each word to portions of words in said association list of **words** and equating non- **matching character strings** ; **rank** ordering candidate lexical rules; and selecting as valid candidate lexical rules which occur more than...

?

File 6:NTIS 1964-2004/Sep W3
(c) 2004 NTIS, Intl Cpyrght All Rights Res

File 2:INSPEC 1969-2004/Sep W2
(c) 2004 Institution of Electrical Engineers

File 8:Ei Compendex(R) 1970-2004/Sep W2
(c) 2004 Elsevier Eng. Info. Inc.

File 256:TecInfoSource 82-2004/Jul
(c)2004 Info.Sources Inc

File 34:SciSearch(R) Cited Ref Sci 1990-2004/Sep W2
(c) 2004 Inst for Sci Info

File 35:Dissertation Abs Online 1861-2004/Aug
(c) 2004 ProQuest Info&Learning

File 65:Inside Conferences 1993-2004/Sep W3
(c) 2004 BLDSC all rts. reserv.

File 94:JICST-EPlus 1985-2004/Aug W4
(c)2004 Japan Science and Tech Corp(JST)

File 95:TEME-Technology & Management 1989-2004/Jun W1
(c) 2004 FIZ TECHNIK

File 99:Wilson Appl. Sci & Tech Abs 1983-2004/Aug
(c) 2004 The HW Wilson Co.

File 111:TGG Natl.Newspaper Index(SM) 1979-2004/Sep 21
(c) 2004 The Gale Group

File 144:Pascal 1973-2004/Sep W2
(c) 2004 INIST/CNRS

File 202:Info. Sci. & Tech. Abs. 1966-2004/Sep 09
(c) 2004 EBSCO Publishing

File 233:Internet & Personal Comp. Abs. 1981-2003/Sep
(c) 2003 EBSCO Pub.

File 266:FEDRIP 2004/Jun
Comp & dist by NTIS, Intl Copyright All Rights Res

File 434:SciSearch(R) Cited Ref Sci 1974-1989/Dec
(c) 1998 Inst for Sci Info

File 438:Library Lit. & Info. Science 1984-2004/Aug
(c) 2004 The HW Wilson Co

File 483:Newspaper Abs Daily 1986-2004/Sep 20
(c) 2004 ProQuest Info&Learning

File 583:Gale Group Globalbase(TM) 1986-2002/Dec 13
(c) 2002 The Gale Group

File 603:Newspaper Abstracts 1984-1988
(c)2001 ProQuest Info&Learning

File 7:Social SciSearch(R) 1972-2004/Sep W2
(c) 2004 Inst for Sci Info

File 121:Brit.Education Index 1976-2004/Q2
(c) 2004 British Education Index

File 142:Social Sciences Abstracts 1983-2004/Aug
(c) 2004 The HW Wilson Co

| Set | Items | Description |
|-----|---------|--|
| S1 | 1171939 | QUERY? OR QUERIE? ? OR SUBQUER? OR SEARCH? OR FETCH? OR RETRIEV? OR TEXTSEARCH? |
| S2 | 769130 | MATCH??? ? |
| S3 | 3440034 | WORD? ? OR TERM? ? |
| S4 | 10434 | CHARACTERSTRING? OR CHARACTER? ?(2N)STRING? ? OR WORDSTEM? OR MORPHEME? OR WORDELEMENT? OR BASETERM? OR BASEWORD? OR LEXEME? |
| S5 | 10561 | S3(2N)(STEM? ? OR ELEMENT? ? OR BASE OR BASES) |
| S6 | 1370834 | SUFFIX? OR PREFIX? OR DERIVATI? OR AFFIX? OR POSTFIX? OR TRUNCAT? OR LEFTTRUNCAT? OR RIGHTTRUNCAT? |
| S7 | 7256868 | RANK? OR RATE OR RATES OR RATED OR RATING? OR SORT??? ? OR SCOR??? ? OR VALUATION? OR TALLY? OR TALLIE? ? OR WEIGH? |
| S8 | 317144 | S7(3N)(RESULT? OR HITLIST? OR REFERENCE? OR RETRIEV? OR HIT |

OR HITS OR OUTPUT? OR OUT()PUT? ? OR RESPONSE? ? OR ANSWER? ?
OR REPLIE? ? OR REPLY?)

S9 93196 S2:S4(3N)S7
S10 31960 S1:S2(5N)S3:S5
S11 807 S10 AND S6
S12 49 S11 AND S8:S9
S13 15 S12/2001:2004
S14 34 S12 NOT S13
S15 27 RD (unique items)

15/7/6 (Item 4 from file: 2)
DIALOG(R)File 2:INSPEC
(c) 2004 Institution of Electrical Engineers. All rts. reserv.

03858708 INSPEC Abstract Number: C91030953

Title: How effective is suffixing ?

Author(s): Harman, D.

Author Affiliation: Lister Hill Center for Biomed. Commun., Nat. Libr. of
Med., Bethesda, MD, USA

Journal: Journal of the American Society for Information Science
vol.42, no.1 p.7-15

Publication Date: Jan. 1991 Country of Publication: USA

CODEN: AISJB6 ISSN: 0002-8231

U.S. Copyright Clearance Center Code: 0002-8231/91/010007-09\$04.00

Language: English Document Type: Journal Paper (JP)

Treatment: Practical (P)

Abstract: The interaction of **suffixing** algorithms and **ranking**
techniques in **retrieval** performance, particularly in an online
environment, was investigated. Three general purpose **suffixing** algorithms
were used for retrieval on the Cranfield 1400, Medlars, and CACM test
collections, with no significant improvement in performance shown for any
of the algorithms. A failure analysis suggested three modifications to
ranking techniques: variable **weighting** of **term** variants, selective
stemming depending on **query** length, and selective stemming depending on
term importance. None of these modifications improved performance.
Recommendations are made regarding the uses of **suffixing** in an online
environment. (17 Refs)

Subfile: C

15/7/20 (Item 1 from file: 202)
DIALOG(R)File 202:Info. Sci. & Tech. Abs.
(c) 2004 EBSCO Publishing. All rts. reserv.

3202143

Beyond Boole: the next logical step.

Author(s): Davis, C H

Corporate Source: Univ. of Illinois, Urbana-Champaign, IL

Bulletin of the American Society for Information Science vol. 21, no. 5
, pages 17-20

Publication Date: Jun-Jul 1995

ISSN: 0095-4403

Language: English

Document Type: Journal Article

Record Type: Abstract

Journal Announcement: 3200

In this article, the author discusses the method of **weighted - term**
searching, which provides **searchers** with a more powerful technique than
Boolean logic. It empowers searchers to control their strategies. From a

systems standpoint, it is easy to implement and represents a straightforward method for getting **ranked output** . It can also be coupled with term **truncation** to provide powerful capabilities for database searching and record display currently unavailable through bibliographic utilities, online search services, or typical database management software packages. The method described can be used profitably in any field by search intermediaries or end users who wish to employ techniques more sophisticated than those afforded by simple Boolean coordination.
? t15/7/26

15/7/26 (Item 2 from file: 233)
DIALOG(R)File 233:Internet & Personal Comp. Abs.
(c) 2003 EBSCO Pub. All rts. reserv.

00336826 94IT01-024

FREESTYLE: LEXIS/NEXIS goes natural

Griffith, Cary

Information Today , January 1, 1994 , v11 n1 p31, 35, 2 Page(s)

ISSN: 8755-6286

Company Name: Mead Data Central; WESTLAW

Product Name: FREESTYLE; LEXIS/NEXIS; WIN

LEGAL LINE column discusses natural language search (NLS) engines, such as that released in November, 1993 by Mead Data Central for its LEXIS/NEXIS service, called FREESTYLE. Calls NLS, or associative retrieval, a revolutionary development in computer-assisted research, which makes online services accessible in ways never before possible, and makes searching as simple as entering a question. Comments on the immaturity of the FREESTYLE product, which was announced on the same day that WESTLAW, a Mead competitor, received product-of-the-year award for its WIN product. Indicates that FREESTYLE is being released gradually, and claims that searches seem to rely on statistical algorithms which examine **queries** , identify relevant search **terms** , **rank** them, and **retrieve** documents which statistically best fit the **query terms** . However, complains that it doesn't automatically **truncate search terms** . Includes three screen displays. (jo)

File 347:JAPIO Nov 1976-2004/May(Updated 040903)
(c) 2004 JPO & JAPIO.
File 350:Derwent WPIX 1963-2004/UD,UM &UP=200459
(c) 2004 Thomson Derwent
File 348:EUROPEAN PATENTS 1978-2004/Sep W02
(c) 2004 European Patent Office.
File 349:PCT FULLTEXT 1979-2002/UB=20040916,UT=20040909
(c) 2004 WIPO/Univentio

| Set | Items | Description |
|-----|---------|--|
| S1 | 124 | AU=ROCHE E? |
| S2 | 29 | AU=SCHABES Y? |
| S3 | 1948461 | QUERY? OR QUERIE? ? OR SUBQUER? OR SEARCH? OR FETCH? OR RETRIEV? OR TEXTSEARCH? OR MATCH? |
| S4 | 125 | S1:S2 |
| S5 | 90860 | S3(5N) (WORD? ? OR TERM? ? OR CHARACTER? OR WORDSTEM? OR MORPHEME? OR WORDELEMENT? OR BASETERM? OR BASEWORD? OR LEXEME?) |
| S6 | 18704 | S3(5N) (CHARACTERSTRING? OR CHARACTER? ? OR STRING? ?) |
| S7 | 11 | S4 AND S5:S6 |

7/9/1 (Item 1 from file: 350)
DIALOG(R)File 350:Derwent WPIX
(c) 2004 Thomson Derwent. All rts. reserv.

014959293 **Image available**
WPI Acc No: 2003-019807/200301
XRPX Acc No: N03-015229

Answering method for a question based on information stored on a computer-readable medium for text indexing and retrieval system, e.g. for retrieving information from World Wide Web

Patent Assignee: GLOBAL INFORMATION RES & TECHNOLOGIES LL (GLOB-N)
Inventor: ROCHE E ; SCHABES Y
Number of Countries: 095 Number of Patents: 002
Patent Family:

| Patent No | Kind | Date | Applicat No | Kind | Date | Week |
|---------------|------|----------|----------------|------|----------|----------|
| WO 200291237 | A1 | 20021114 | WO 2001US14708 | A | 20010507 | 200301 B |
| AU 2001259591 | A1 | 20021118 | AU 2001259591 | A | 20010507 | 200452 |
| | | | WO 2001US14708 | A | 20010507 | |

Priority Applications (No Type Date): WO 2001US14708 A 20010507

Patent Details:

| Patent No | Kind | Lan | Pg | Main IPC | Filing Notes |
|-----------|------|-----|----|----------|--------------|
|-----------|------|-----|----|----------|--------------|

| | | | | | |
|--------------|----|---|-----|-------------|--|
| WO 200291237 | A1 | E | 126 | G06F-017/30 | |
|--------------|----|---|-----|-------------|--|

Designated States (National): AE AG AL AM AT AU AZ BA BB BG BR BY BZ CA CH CN CO CR CU CZ DE DK DM DZ EC EE ES FI GB GD GE GH GM HR HU ID IL IN IS JP KE KG KP KR KZ LC LK LR LS LT LU LV MA MD MG MK MN MW MX MZ NO NZ PL PT RO RU SD SE SG SI SK SL TJ TM TR TT TZ UA UG UZ VN YU ZA ZW

Designated States (Regional): AT BE CH CY DE DK EA ES FI FR GB GH GM GR IE IT KE LS LU MC MW MZ NL OA PT SD SE SL SZ TR TZ UG ZW

| | | | | | |
|---------------|----|--|--|-------------|------------------------------|
| AU 2001259591 | A1 | | | G06F-017/30 | Based on patent WO 200291237 |
|---------------|----|--|--|-------------|------------------------------|

Abstract (Basic): WO 200291237 A1

NOVELTY - The method involves receiving a question. The question is parsed to obtain an analyzed question. The analyzed question is matched to a set of predetermined question patterns to obtain matched question patterns. The matched question patterns are transformed into one or more partially unspecified statements. Each of the partially unspecified statements is missing a portion corresponding to an answer. Partially unspecified queries are generated corresponding to the partially unspecified statements. Answers are obtained by matching the partially unspecified queries to stored information

DETAILED DESCRIPTION - INDEPENDENT CLAIMS are also included for the following:

- (a) an apparatus for answering a natural language question;
- (b) computer-executable process steps stored on a computer-readable medium.

USE - For processing natural language question.

ADVANTAGE - Allows matching to be conducted without strict ordering of **query terms**.

DESCRIPTION OF DRAWING(S) - The figure shows the overall process of obtaining an answer or answers for a natural language question.

pp; 126 DwgNo 2/5

Title Terms: ANSWER; METHOD; QUESTION; BASED; INFORMATION; STORAGE; COMPUTER; READ; MEDIUM; TEXT; INDEX; RETRIEVAL; SYSTEM; RETRIEVAL; INFORMATION; WORLD; WIDE; WEB

Derwent Class: T01

International Patent Class (Main): G06F-017/30

File Segment: EPI

Manual Codes (EPI/S-X): T01-J05B1; T01-J16C3; T01-N03A2; T01-S03

7/9/2 (Item 2 from file: 350)

DIALOG(R)File 350:Derwent WPIX

(c) 2004 Thomson Derwent. All rts. reserv.

014699005 **Image available**

WPI Acc No: 2002-519709/200255

Related WPI Acc No: 2002-106021

XRPX Acc No: N02-411350

Information search and retrieval method through Internet involves locating matches for required information within identified contexts containing partially unspecified terms

Patent Assignee: GLOBAL INFORMATION RES & TECHNOLOGIES LL (GLOB-N); ROCHE E (ROCH-I); SCHABES Y (SCHA-I)

Inventor: ROCHE E ; SCHABES Y

Number of Countries: 097 Number of Patents: 003

Patent Family:

| Patent No | Kind | Date | Applicat No | Kind | Date | Week |
|----------------|------|----------|----------------|------|----------|----------|
| WO 200246970 | A2 | 20020613 | WO 2001US46542 | A | 20011205 | 200255 B |
| US 20020123994 | A1 | 20020905 | US 2000559223 | A | 20000426 | 200260 |
| | | | US 2000251608 | A | 20001205 | |
| | | | US 20014952 | A | 20011205 | |
| AU 200220219 | A | 20020618 | AU 200220219 | A | 20011205 | 200262 |

Priority Applications (No Type Date): US 20014952 A 20011205; US 2000251608 P 20001205; US 2000559223 A 20000426

Patent Details:

Patent No Kind Lan Pg Main IPC Filing Notes

WO 200246970 A2 E 98 G06F-017/30

Designated States (National): AE AG AL AM AT AU AZ BA BB BG BR BY BZ CA CH CN CO CR CU CZ DE DK DM DZ EC EE ES FI GB GD GE GH GM HR HU ID IL IN IS JP KE KG KP KR KZ LC LK LR LS LT LU LV MA MD MG MK MN MW MX MZ NO NZ PL PT RO RU SD SE SG SI SK SL TJ TM TR TT TZ UA UG UZ VN YU ZA ZW

Designated States (Regional): AT BE CH CY DE DK EA ES FI FR GB GH GM GR IE IT KE LS LU MC MW MZ NL OA PT SD SE SL SZ TR TZ UG ZM ZW

US 20020123994 A1 G06F-007/00 CIP of application US 2000559223

Provisional application US 2000251608

AU 200220219 A G06F-017/30 Based on patent WO 200246970

Abstract (Basic): WO 200246970 A2

NOVELTY - Contexts in an index that contain fully specified **terms** in a **query** and partially unspecified **terms** of required information,

are identified. **Matches** for the required information are located within the identified contexts.

DETAILED DESCRIPTION - INDEPENDENT CLAIMS are included for the following:

(1) Apparatus for searching and retrieving required information; and

(2) Computer program for searching and retrieving required information.

USE - For searching and retrieving information from web page of Internet.

ADVANTAGE - Allows for identification of **matches** in documents in the **matching terms** need not appear in the same relative order as in the query and in which there are intervening **words** between the **matching terms**. A **search query** is processed efficiently and fast.

DESCRIPTION OF DRAWING(S) - The figure shows the flowchart of the extended matching process.

pp; 98 DwgNo 18/27

Title Terms: INFORMATION; SEARCH; RETRIEVAL; METHOD; THROUGH; LOCATE; MATCH ; REQUIRE; INFORMATION; IDENTIFY; CONTEXT; CONTAIN; TERM

Derwent Class: T01

International Patent Class (Main): G06F-007/00; G06F-017/30

File Segment: EPI

Manual Codes (EPI/S-X): T01-J05B1; T01-N03A2; T01-S03

7/9/3 (Item 3 from file: 350)

DIALOG(R)File 350:Derwent WPIX

(c) 2004 Thomson Derwent. All rts. reserv.

012900837 **Image available**

WPI Acc No: 2000-072673/200006

XPX Acc No: N00-056840

Misspelled words correcting method in machine translation system, word processing system, etc

Patent Assignee: TERAGRAM CORP (TERA-N); ROCHE E (ROCH-I); SCHABES Y (SCHA-I); GLOBAL INFORMATION RES & TECHNOLOGIES LL (GLOB-N)

Inventor: **ROCHE E ; SCHABES Y**

Number of Countries: 083 Number of Patents: 005

Patent Family:

| Patent No | Kind | Date | Applicat No | Kind | Date | Week |
|----------------|------|----------|---------------|------|----------|----------|
| WO 9962000 | A2 | 19991202 | WO 99US11713 | A | 19990526 | 200006 B |
| AU 9941003 | A | 19991213 | AU 9941003 | A | 19990526 | 200020 |
| EP 1145141 | A2 | 20011017 | EP 99924524 | A | 19990526 | 200169 |
| | | | WO 99US11713 | A | 19990526 | |
| US 6424983 | B1 | 20020723 | US 9884535 | A | 19980526 | 200254 |
| US 20040093567 | A1 | 20040513 | US 9884535 | A | 19980526 | 200432 |
| | | | US 2002153460 | A | 20020522 | |

Priority Applications (No Type Date): US 9884535 A 19980526; US 2002153460 A 20020522

Patent Details:

Patent No Kind Lan Pg Main IPC Filing Notes

WO 9962000 A2 E 145 G06F-017/27

Designated States (National): AL AM AT AU AZ BA BB BG BR BY CA CH CN CU CZ DE DK EE ES FI GB GE GH GM HR HU ID IL IS JP KE KG KP KR KZ LC LK LR LS LT LU LV MD MG MK MN MW MX NO NZ PL PT RO RU SD SE SG SI SK SL TJ TM TR TT UA UG UZ VN YU ZW

Designated States (Regional): AT BE CH CY DE DK EA ES FI FR GB GH GM GR IE IT KE LS LU MC MW NL OA PT SD SE SL SZ UG ZW

AU 9941003 A G06F-017/27 Based on patent WO 9962000
EP 1145141 A2 E G06F-017/27 Based on patent WO 9962000
Designated States (Regional): AT BE CH CY DE DK ES FI FR GB GR IE IT LI
LU MC NL PT SE
US 6424983 B1 G06F-017/00
US 20040093567 A1 G06F-017/00 Cont of application US 9884535
Cont of patent US 6424983

Abstract (Basic): WO 9962000 A2

NOVELTY - Misspelled word in an input text is detected by comparing each word of text to a dictionary database. Then, a list of alternative words for the misspelled word is determined. Alternative words are then ranked based on the context of the input text. The misspelled word in text is replaced with one alternative word selected from the list.

DETAILED DESCRIPTION - The word is detected as misspelled when the word either does not match with any words in the dictionary database or the word spelled correctly but corresponds to one of the several words which are substantially similar. Each of lexicon finite state machine (FSM) including plural reference words and phonetic representation of each reference word, is stored. An input FSM including misspelled words and its phonetic representation is generated. One or more reference words from the lexicon FSMs is selected based on the input FSM, one or more reference words corresponding to either spelling or phonetic representation of the misspelled word. Then the selected reference words are added to the list of alternative words. INDEPENDENT CLAIMS are also included for the following:

- (a) word processing method;
- (b) optical character recognition method;
- (c) machine translation method;
- (d) computer readable memory storing computer executable process for correcting misspelled words in input text;
- (e) apparatus for retrieving text from source

USE - For correcting misspelled words, incorrectly used words in input text in machine translation system, word processing system and text indexing and retrieval system such as world wide web search engine.

ADVANTAGE - Enables to correct improper use of commonly confused words and also the words that are spelled correctly but that are improper in context.

DESCRIPTION OF DRAWING(S) - The figure shows the operation of spelling and grammar checking system.

pp; 145 DwgNo 3/25

Title Terms: WORD; CORRECT; METHOD; MACHINE; TRANSLATION; SYSTEM; WORD; PROCESS; SYSTEM

Derwent Class: T01

International Patent Class (Main): G06F-017/00; G06F-017/27

International Patent Class (Additional): G06F-017/21; G06F-017/24

File Segment: EPI

Manual Codes (EPI/S-X): T01-J05B4; T01-J11A1; T01-J14; T01-S03

7/9/4 (Item 4 from file: 350)

DIALOG(R)File 350:Derwent WPIX

(c) 2004 Thomson Derwent. All rts. reserv.

010680102 **Image available**

WPI Acc No: 1996-177057/199618

XRPX Acc No: N96-148753

Context tag for speech recognition - uses definite limited state transducers to allocate part of speech tag to input sentence such that

allocated speech tag matches with required word
Patent Assignee: MITSUBISHI ELECTRIC CORP (MITQ); MITSUBISHI ELECTRIC
INFORMATION TECHNOLO (MITQ)

Inventor: **ROCHE E ; SCHABES Y**

Number of Countries: 002 Number of Patents: 002

Patent Family:

| Patent No | Kind | Date | Applicat No | Kind | Date | Week |
|------------|------|----------|-------------|------|----------|----------|
| JP 8055122 | A | 19960227 | JP 95157872 | A | 19950623 | 199618 B |
| US 5610812 | A | 19970311 | US 94264981 | A | 19940624 | 199716 |

Priority Applications (No Type Date): US 94264981 A 19940624

Patent Details:

| Patent No | Kind | Lan Pg | Main IPC | Filing Notes |
|------------|------|--------|-------------|--------------|
| JP 8055122 | A | 14 | G06F-017/27 | |
| US 5610812 | A | 88 | G06F-007/27 | |

Abstract (Basic): JP 8055122 A

The context tag has a sequence of context rules, which is converted into indefinite parts using a transducer. A photo-composing machine and a data machine are connected to the main part. A definite limited states transducer converts the indefinite parts into definite limited states.

The portion of a speech tag which almost **matches** with the required **word** is added to the input sentence using the transducer. The other portion of the speech tag are omitted as they are indefinite. The time required to select the portion of the speech tag is proportional to the number of words in the input sentence.

ADVANTAGE - Gives high accuracy. Enables high speed operation. Operates independent of number of rules. Enables high speed grammatical check, spelling check, information extraction, optical character recognition.

Dwg.1/10

Abstract (Equivalent): US 5610812 A

A computer system for correcting part of speech tags of words of sentences in a text, comprising:
means for receiving an initially tagged input sentence; and,
a contextual part of speech tagger for correcting part-of-speech tags of the words of said initially tagged input sentence, said tagger including a deterministic finite state transducer for tagging said words in accordance with context and in a single pass.

Dwg.8/13

Title Terms: CONTEXT; TAG; SPEECH; RECOGNISE; DEFINITE; LIMIT; STATE; TRANSDUCER; ALLOCATE; PART; SPEECH; TAG; INPUT; SENTENCE; ALLOCATE; SPEECH; TAG; MATCH; REQUIRE; WORD

Derwent Class: T01; W04

International Patent Class (Main): G06F-007/27; G06F-017/27

File Segment: EPI

Manual Codes (EPI/S-X): T01-J11A; W04-V04A; W04-V05C

7/5/5 (Item 1 from file: 349)

DIALOG(R)File 349:PCT FULLTEXT

(c) 2004 WIPO/Univentio. All rts. reserv.

00957081 **Image available**

SYSTEM FOR ANSWERING NATURAL LANGUAGE QUESTIONS

SYSTEME PERMETTANT DE REPONDRE A DES QUESTIONS EN LANGAGE NATUREL

Patent Applicant/Assignee:

GLOBAL INFORMATION RESEARCH AND TECHNOLOGIES LLC, 236 Huntington Avenue,
Boston, MA 02115-4701, US, US (Residence), US (Nationality)

Inventor(s):

SCHABES Yves , c/o Teragram, 236 Huntington Avenue, Boston, MA 02115, US

ROCHE Emmanuel , c/o Teragram, 236 Huntington Avenue, Boston, MA 02115,
US

Legal Representative:

PASTERNAK Sam (agent), Choate, Hall & Stewart, 53 State Street, Exchange
Place, Boston, MA 02109, US,

Patent and Priority Information (Country, Number, Date):

Patent: WO 200291237 A1 20021114 (WO 0291237)

Application: WO 2001US14708 20010507 (PCT/WO US0114708)

Priority Application: WO 2001US14708 20010507

Designated States:

(Protection type is "patent" unless otherwise stated - for applications
prior to 2004)

AE AG AL AM AT AU AZ BA BB BG BR BY BZ CA CH CN CO CR CU CZ DE DK DM DZ
EC EE ES FI GB GD GE GH GM HR HU ID IL IN IS JP KE KG KP KR KZ LC LK LR
LS LT LU LV MA MD MG MK MN MW MX MZ NO NZ PL PT RO RU SD SE SG SI SK SL
TJ TM TR TT TZ UA UG UZ VN YU ZA ZW

(EP) AT BE CH CY DE DK ES FI FR GB GR IE IT LU MC NL PT SE TR

(OA) BF BJ CF CG CI CM GA GN GW ML MR NE SN TD TG

(AP) GH GM KE LS MW MZ SD SL SZ TZ UG ZW

(EA) AM AZ BY KG KZ MD RU TJ TM

Main International Patent Class: G06F-017/30

Publication Language: English

Filing Language: English

Fulltext Availability:

Detailed Description

Claims

Fulltext Word Count: 28924

English Abstract

The present invention is a system for answering a natural language questions. The system receives a question and transforms the question into one or more partially unspecified queries. The system then identifies matches for the queries in a body of information. The matches are optionally ranked, preferably based on the number of times each match is identified. The matches are provided as answers to the questions.

French Abstract

La presente invention concerne un systeme permettant de repondre a des questions en langage naturel. Ce systeme recoit une question et transforme cette question en une ou plusieurs demandes partiellement non specifiees. Ce systeme identifie ensuite des correspondances de ces demandes dans un corps d'informations. Ces correspondances sont eventuellement classees, de preference en fonction du nombre d'identifications de chaque correspondance. Ces correspondances sont fournies comme reponses aux questions.

Legal Status (Type, Date, Text)

Publication 20021114 A1 With international search report.

7/5/6 (Item 2 from file: 349)

DIALOG(R)File 349:PCT FULLTEXT

(c) 2004 WIPO/Univentio. All rts. reserv.

00912809 **Image available**

SYSTEM FOR FULFILLING AN INFORMATION NEED USING EXTENDED MATCHING
TECHNIQUES

SYSTEME PERMETTANT DE REpondre A UN BESOIN D'INFORMATION PAR DES TECHNIQUES
D'APPARIEMENT APPROFONDIES

Patent Applicant/Assignee:

GLOBAL INFORMATION RESEARCH AND TECHNOLOGIES LLC, 236 Huntington Avenue,

Boston, MA 02115-4701, US, US (Residence), US (Nationality)
Inventor(s):
SCHABES Yves , c/o Teragram, 236 Huntington Avenue, Boston, MA 02115, US
ROCHE Emmanuel , c/o Teragram, 236 Huntington Avenue, Boston, MA 02115, US

Legal Representative:

HAMILTON John A (agent), Choate, Hall & Stewart, Exchange Place, 53 State Street, Boston, MA 02109, US,

Patent and Priority Information (Country, Number, Date):

Patent: WO 200246970 A2-A3 20020613 (WO 0246970)
Application: WO 2001US46542 20011205 (PCT/WO US01046542)
Priority Application: US 2000251608 20001205

Designated States:

(Protection type is "patent" unless otherwise stated - for applications prior to 2004)

AE AG AL AM AT AU AZ BA BB BG BR BY BZ CA CH CN CO CR CU CZ DE DK DM DZ
EC EE ES FI GB GD GE GH GM HR HU ID IL IN IS JP KE KG KP KR KZ LC LK LR
LS LT LU LV MA MD MG MK MN MW MX MZ NO NZ PL PT RO RU SD SE SG SI SK SL
TJ TM TR TT TZ UA UG UZ VN YU ZA ZW

(EP) AT BE CH CY DE DK ES FI FR GB GR IE IT LU MC NL PT SE TR

(OA) BF BJ CF CG CI CM GA GN GQ GW ML MR NE SN TD TG

(AP) GH GM KE LS MW MZ SD SL SZ TZ UG ZM ZW

(EA) AM AZ BY KG KZ MD RU TJ TM

Main International Patent Class: G06F-017/30

Publication Language: English

Filing Language: English

Fulltext Availability:

Detailed Description

Claims

Fulltext Word Count: 24863

English Abstract

The invention offers new approaches to fulfilling an information need, in particular to finding a result for a query based on a large body of information such as a collection of documents. The invention accepts a query containing an unspecified portion that expresses the information need. The invention locates matches for the query within a body of information and returns the matches or portions thereof in addition to or instead of identifiers for documents in which the matches are found. The invention allows placement of term ordering restrictions, and allows intervening words between the search terms as they appear in the searched documents or contexts. The invention ranks the matches in order to provide the most relevant information. One preferred method of ranking considers the number of instances of a match among a plurality of documents. The invention further defines a new type of index that includes contexts in which terms occur and provides methods of searching such indices to fulfill an information need.

French Abstract

L'invention presente de nouvelles approches permettant de repondre a un besoin d'information, en particulier de trouver un resultat a une interrogation en fonction d'un grand nombre d'informations, tel qu'une collection de documents. Selon l'invention, le systeme accepte une interrogation contenant une partie non specifiee qui exprime le besoin d'information. Ce systeme localise des correspondances pour cette interrogation dans un corps d'informations, et renvoie ces correspondances, ou des parties de celles-ci, en plus ou a la place d'identificateurs de documents dans lesquels on trouve ces correspondances. L'invention permet de placer des restrictions de classement de termes, et elle permet de faire intervenir des mots entre

les termes de recherche, a mesure qu'ils apparaissent dans les documents ou les contextes explores. L'invention classe les correspondances dans l'ordre pour fournir l'information la plus pertinente. Dans un procede de classement prefere, le nombre d'exemples de correspondance parmi une pluralite de documents est pris en consideration. Le systeme definit egalement un nouveau type d'index qui comporte des contextes dans lesquels des termes apparaissent, et met a disposition des procedes de recherche de tels indices pour repondre a un besoin d'information.

Legal Status (Type, Date, Text)

Publication 20020613 A2 Without international search report and to be republished upon receipt of that report.

Examination 20030206 Request for preliminary examination prior to end of 19th month from priority date

Search Rpt 20040226 Late publication of international search report

Republication 20040226 A3 With international search report.

7/5/7 (Item 3 from file: 349)

DIALOG(R)File 349:PCT FULLTEXT

(c) 2004 WIPO/Univentio. All rts. reserv.

00851704 **Image available**

SYSTEM FOR ANSWERING NATURAL LANGUAGE QUESTIONS

SYSTEME POUR REPONDRE A DES QUESTIONS FORMULEES EN LANGAGE NATUREL

Patent Applicant/Assignee:

GLOBAL INFORMATION RESEARCH AND TECHNOLOGIES LLC, 236 Huntington Avenue, Boston, MA 02115-4701, US, US (Residence), US (Nationality)

Inventor(s):

SCHABES Yves , c/o Teragram, 236 Huntington Avenue, Boston, MA 02115, US

ROCHE Emmanuel , c/o Teragram, 236 Huntington Avenue, Boston, MA 02115, US

Legal Representative:

PASTERNAK Sam (agent), Choate, Hall & Stewart, 53 State Street, Exchange Place, Boston, MA 02109, US,

Patent and Priority Information (Country, Number, Date):

Patent: WO 200184376 A2-A3 20011108 (WO 0184376)

Application: WO 2001US13935 20010430 (PCT/WO US0113935)

Priority Application: US 2000200766 20000428

Designated States:

(Protection type is "patent" unless otherwise stated - for applications prior to 2004)

AE AG AL AM AT AU AZ BA BB BG BR BY BZ CA CH CN CO CR CU CZ DE DK DM DZ
EE ES FI GB GD GE GH GM HR HU ID IL IN IS JP KE KG KP KR KZ LC LK LR LS
LT LU LV MA MD MG MK MN MW MX MZ NO NZ PL PT RO RU SD SE SG SI SK SL TJ
TM TR TT TZ UA UG UZ VN YU ZA ZW

(EP) AT BE CH CY DE DK ES FI FR GB GR IE IT LU MC NL PT SE TR

(OA) BF BJ CF CG CI CM GA GN GW ML MR NE SN TD TG

(AP) GH GM KE LS MW MZ SD SL SZ TZ UG ZW

(EA) AM AZ BY KG KZ MD RU TJ TM

Main International Patent Class: G06F-017/30

Publication Language: English

Filing Language: English

Fulltext Availability:

Detailed Description

Claims

Fulltext Word Count: 29154

English Abstract

The present invention is a system for answering a natural language

question. The system receives a question and transforms the question into one or more partially unspecified queries. The system then identifies matches for the queries in a body of information. The matches are optionally ranked, preferably based on the number of times each match is identified. The matches are provided as answers to the questions.

French Abstract

La presente invention concerne un systeme qui permet de repondre a une question formulee en langage naturel. Le systeme recoit une question et la transforme en une ou plusieurs interrogations partiellement non precisees. Le systeme identifie ensuite des equivalences pour ces interrogations dans un corps de donnees. Les equivalences sont facultativement classees par ordre, de preference sur la base de la frequence d'identification de chaque equivalence. Les equivalences sont fournies comme reponses aux questions.

Legal Status (Type, Date, Text)

Publication 20011108 A2 Without international search report and to be republished upon receipt of that report.
Examination 20020328 Request for preliminary examination prior to end of 19th month from priority date
Search Rpt 20020725 Late publication of international search report
Republication 20020725 A3 With international search report.
Republication 20020725 A3 Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.

7/5/8 (Item 4 from file: 349)

DIALOG(R)File 349:PCT FULLTEXT

(c) 2004 WIPO/Univentio. All rts. reserv.

00848502 **Image available**

SYSTEM FOR FULFILLING AN INFORMATION NEED

SYSTEME REPONDANT A UN BESOIN D'INFORMATION

Patent Applicant/Assignee:

GLOBAL INFORMATION RESEARCH AND TECHNOLOGIES LLC, c/o Teragram Corp., 236 Huntington Avenue, Boston, MA 02115-4701, US, US (Residence), US (Nationality)

Inventor(s):

ROCHE Emmanuel , Gobal Information Research and Technologies LLC, c/o Teragram, 236 Huntington Avenue, Boston, MA 02215-4701, US,

SCHABES Yves , Global Information Research and Technologies LLC, c/o Teragram, 236 Huntington Avenue, Boston, MA 02115-4701, US

Legal Representative:

GERBER Monica R (agent), Choate, Hall & Stewart, Exchange Place, 53 State Street, Boston, MA 02109, US,

Patent and Priority Information (Country, Number, Date):

Patent: WO 200182114 A2-A3 20011101 (WO 0182114)

Application: WO 2001US13150 20010424 (PCT/WO US0113150)

Priority Application: US 2000559223 20000426

Designated States:

(Protection type is "patent" unless otherwise stated - for applications prior to 2004)

AE AG AL AM AT AU AZ BA BB BG BR BY BZ CA CH CN CO CR CU CZ DE DK DM DZ
EE ES FI GB GD GE GH GM HR HU ID IL IN IS JP KE KG KP KR KZ LC LK LR LS
LT LU LV MA MD MG MK MN MW MX MZ NO NZ PL PT RO RU SD SE SG SI SK SL TJ
TM TR TT TZ UA UG UZ VN YU ZA ZW

(EP) AT BE CH CY DE DK ES FI FR GB GR IE IT LU MC NL PT SE TR

(OA) BF BJ CF CG CI CM GA GN GW ML MR NE SN TD TG

(AP) GH GM KE LS MW MZ SD SL SZ TZ UG ZW

(EA) AM AZ BY KG KZ MD RU TJ TM
Main International Patent Class: G06F-017/30
Publication Language: English
Filing Language: English
Fulltext Availability:
 Detailed Description
 Claims
Fulltext Word Count: 26836

English Abstract

The invention offers new approaches to fulfilling an information need, in particular to finding a result for a query based on a large body of information such as a collection of documents. The invention accepts a query containing an unspecified portion that expresses the information need. The invention locates matches for the query within a body of information and returns the matches or portions thereof in addition to or instead of identifiers for documents in which the matches are found. The invention ranks the matches in order to provide the most relevant information. One preferred method of ranking considers the number of instances of a match among a plurality of documents. The invention further defines a new type of index that includes contexts in which **terms** occur and provides methods of **searching** such indices to fulfill an information need.

French Abstract

L'invention concerne de nouvelles approches de reponse a un besoin d'information permettant, en particulier, de trouver une reponse a une demande fondee sur une grande quantite d'informations, telles qu'une collection de documents. L'invention permet d'accepter une demande contenant une partie non specifiee exprimant un besoin information. Elle permet de localiser des correspondances entre la demande et un corps d'informations, et renvoie les correspondances ou des parties de celles-ci avec des identificateurs de documents dans lesquels des correspondances ont ete trouvees ou a la place de ceux-ci. Elle permet de classer les correspondances de facon a fournir les informations les plus pertinentes. Un procede de classement prefere, considere le nombre d'instances d'une correspondance dans plusieurs documents. Elle permet egalement de definir un nouveau type d'index comprenant des contextes dans lesquels des termes apparaissent, et fournit des procedes de recherche d'indices permettant de repondre a un besoin d'information.

Legal Status (Type, Date, Text)

Publication 20011101 A2 Without international search report and to be republished upon receipt of that report.
Examination 20020321 Request for preliminary examination prior to end of 19th month from priority date
Search Rpt 20030912 Late publication of international search report
Republication 20030912 A3 With international search report.
Republication 20030912 A3 Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.

7/5/9 (Item 5 from file: 349)
DIALOG(R)File 349:PCT FULLTEXT
(c) 2004 WIPO/Univentio. All rts. reserv.

00571517 **Image available**
TEXT TRANSLATION SYSTEM
SYSTEME DE TRADUCTION DE TEXTES
Patent Applicant/Assignee:

GLOBAL INFORMATION RESEARCH AND TECHNOLOGIES LLC,
Inventor(s):

ROCHE Emmanuel ,
SCHABES Yves

Patent and Priority Information (Country, Number, Date):

Patent: WO 200034890 A2 20000615 (WO 0034890)

Application: WO 99US29273 19991210 (PCT/WO US9929273)

Priority Application: US 98111702 19981210

Designated States:

(Protection type is "patent" unless otherwise stated - for applications prior to 2004)

AE AL AM AT AU AZ BA BB BG BR BY CA CH CN CU CZ DE DK EE ES FI GB GD GE
GH GM HR HU ID IL IN IS JP KE KG KP KR KZ LC LK LR LS LT LU LV MD MG MK
MN MW MX NO NZ PL PT RO RU SD SE SG SI SK SL TJ TM TR TT UA UG UZ VN YU
ZA ZW GH GM KE LS MW SD SL SZ TZ UG ZW AM AZ BY KG KZ MD RU TJ TM AT BE
CH CY DE DK ES FI FR GB GR IE IT LU MC NL PT SE BF BJ CF CG CI CM GA GN
GW ML MR NE SN TD TG

Main International Patent Class: G06F-017/28

Publication Language: English

Fulltext Availability:

Detailed Description

Claims

Fulltext Word Count: 8324

English Abstract

A system builds a text fragment database for use in translating fragments of text from a source language into a target language. The system first stores a sentence database in memory, the sentence database comprising a plurality of sentence pairs, each sentence pair including a sentence in the source language and a corresponding sentence in the target language. The system then locates corresponding source and target text fragments in corresponding source and target language sentences, respectively, and stores the source text fragment together with the target text fragment in the text fragment database. The text fragment database can then be used to translate text from the source language into the target language. To this end, the system inputs text in the source language, extracts a text fragment from the input text, and locates the extracted text fragment in the text fragment database. The system then retrieves, from the text fragment database, a text fragment in the target language that corresponds to the extracted text fragment, and outputs the retrieved text fragment.

French Abstract

L'invention concerne un systeme qui constitue une base de donnees de fragments de texte pour la traduction de fragments de texte d'une langue source vers une langue cible. Tout d'abord, le systeme met en memoire une base de donnees de phrases, laquelle comporte une pluralite de paires de phrases, chaque paire se composant d'une phrase dans la langue source et d'une phrase correspondante dans la langue cible. Le systeme repere ensuite les fragments de texte source et cible correspondants dans les phrases correspondantes en langue source et en langue cible respectivement, et met en memoire le fragment de texte source associe au fragment de texte cible dans la base de donnees de fragments de texte. La base de donnees de fragments de texte peut ensuite etre utilisee pour la traduction de textes de la langue source vers la langue cible. A cet effet, le systeme introduit du texte en langue source, extrait un fragment du texte introduit, et repere le fragment de texte extrait dans la base de donnees de fragments de texte. Ce systeme recupere ensuite, de la base de donnees de fragments de texte, un fragment en langue cible, qui correspond au fragment extrait, et produit le fragment de texte recupere.

7/5/10 (Item 6 from file: 349)
DIALOG(R)File 349:PCT FULLTEXT
(c) 2004 WIPO/Univentio. All rts. reserv.

00545199 **Image available**

SEARCH AND INDEX HOSTING SYSTEM

SYSTEME DE RECHERCHE ET D'HEBERGEMENT D'INDEX

Patent Applicant/Assignee:

GLOBAL INFORMATION RESEARCH AND TECHNOLOGIES LLC,

Inventor(s):

SCHABES Yves ,

ROCHE Emmanuel ,

BROWN Ryan

Patent and Priority Information (Country, Number, Date):

Patent: WO 200008572 A1 20000217 (WO 0008572)

Application: WO 99US17359 19990802 (PCT/WO US9917359)

Priority Application: US 98130420 19980806

Designated States:

(Protection type is "patent" unless otherwise stated - for applications prior to 2004)

AE AL AM AT AU AZ BA BB BG BR BY CA CH CN CU CZ DE DK EE ES FI GB GD GE
GH GM HR HU ID IL IN IS JP KE KG KP KR KZ LC LK LR LS LT LU LV MD MG MK
MN MW MX NO NZ PL PT RO RU SD SE SG SI SK SL TJ TM TR TT UA UG UZ VN YU
ZA ZW GH GM KE LS MW SD SL SZ UG ZW AM AZ BY KG KZ MD RU TJ TM AT BE CH
CY DE DK ES FI FR GB GR IE IT LU MC NL PT SE BF BJ CF CG CI CM GA GN GW
ML MR NE SN TD TG

Main International Patent Class: G06F-017/30

Publication Language: English

Fulltext Availability:

Detailed Description

Claims

Fulltext Word Count: 12008

English Abstract

The system initiates a search at a first network site for user-specified data in a remote database at a second network site and conducts the search at a third network site (e.g., at a host computer's site). To begin, the system receives, at the first network site, a provider identifier associated with the database from the second network site. Thereafter, the user-specified data is input at the first network site, following which the user-specified data and the provider identifier are output from the first network site to the third network site. The system then searches for the user-specified data in a database at the third network site using the provider identifier. This database at the third network site includes data that corresponds to data stored in the remote database at the second network site.

French Abstract

Le systeme lance sur un premier site de reseau une recherche de donnees propres a un utilisateur dans une base de donnees eloignee d'un deuxieme site de reseau, puis effectue la recherche sur un troisieme site de reseau (par exemple celui d'un ordinateur hote). Au commencement, le systeme recoit au premier site de reseau l'identification d'un prestataire associee a une base de donnees du deuxieme site de reseau. Puis, les donnees propres a l'utilisateur sont introduites dans le premier site de reseau et transmises, avec l'identificateur de prestataire du premier site de reseau au troisieme site de reseau. Le systeme recherche alors les donnees propres a l'utilisateur dans la base de donnees du troisieme site de reseau a l'aide de l'identificateur de prestataire. Ladite base de donnees comprend des donnees correspondant

aux donnees stockees dans la base de donnees distante du deuxieme site de reseau.

7/5/11 (Item 7 from file: 349)
DIALOG(R)File 349:PCT FULLTEXT
(c) 2004 WIPO/Univentio. All rts. reserv.

00530648 **Image available**

SPELLING AND GRAMMAR CHECKING SYSTEM

SYSTEME DE VERIFICATION ORTHOGRAPHIQUE ET GRAMMATICALE

Patent Applicant/Assignee:

TERAGRAM CORPORATION, 236 Huntington Avenue, Boston, MA 02115-4701, US,
US (Residence), US (Nationality)

Inventor(s):

SCHABES Yves , 770 Boylston Street #12B, Boston, MA 02199, US,

ROCHE Emmanuel , 33 Clarendon Road, Belmont, MA 02178, US

Legal Representative:

NUGENT Elizabeth E (agent), Choate, Hall & Stewart, Exchange Place, 53
State Street, Boston, MA 02109, US,

Patent and Priority Information (Country, Number, Date):

Patent: WO 9962000 A2-A3 19991202

Application: WO 99US11713 19990526 (PCT/WO US9911713)

Priority Application: US 9884535 19980526

Designated States:

(Protection type is "patent" unless otherwise stated - for applications prior to 2004)

AL AM AT AU AZ BA BB BG BR BY CA CH CN CU CZ DE DK EE ES FI GB GE GH GM

HR HU ID IL IS JP KE KG KP KR KZ LC LK LR LS LT LU LV MD MG MK MN MW MX

NO NZ PL PT RO RU SD SE SG SI SK SL TJ TM TR TT UA UG UZ VN YU ZW

(EP) AT BE CH CY DE DK ES FI FR GB GR IE IT LU MC NL PT SE

(OA) BF BJ CF CG CI CM GA GN GW ML MR NE SN TD TG

(AP) GH GM KE LS MW SD SL SZ UG ZW

(EA) AM AZ BY KG KZ MD RU TJ TM

Main International Patent Class: G06F-017/27

International Patent Class: G06F-017/28; G06F-017/30

Publication Language: English

Filing Language: English

Fulltext Availability:

Detailed Description

Claims

Fulltext Word Count: 26445

English Abstract

A system of correcting misspelled words in input text detects a misspelled word in the input text, determines a list of alternative words for the misspelled word, and ranks the list of alternative words based on a context of the input text. The system then selects one of the alternative words from the list, and replaces the misspelled word in the text with the selected one of the alternative words.

French Abstract

Ce systeme, permettant de corriger des mots mal orthographies dans un texte d'entree, detecte le mot mal orthographie, arrete une liste de mots de remplacement et classe cette liste de mots en fonction du contexte du texte d'entree. Le systeme selectionne l'un des mots de remplacement qui va remplacer le mot mal orthographie.

Legal Status (Type, Date, Text)

Rev Srch Rpt 20011115 Late publication of revised international search report

Search Rpt 20010607 Late publication of international search report

Republication 20011115 A3 With international search report.

?